JMIR Serious Games

Contents

Original Papers

<u>Original Paper</u>

A Game to Deal With Alcohol Abuse (Jib): Development and Game Experience Evaluation

Darlinton Barbosa Feres Carvalho[1*], BSc, MSc, PhD; Daniel Bueno Domingueti[1*], BSc; Sandro Martins De Almeida Santos[2*], BSc, MSc, PhD; Diego Roberto Colombo Dias[1*], BSc, MSc, PhD

[1]Federal University of São João del-Rei, Departament of Computer Science, São João del-Rei, Brazil

[2]Federal University of Amazonas, Department of Anthropology, Manaus, Brazil

[*]all authors contributed equally

Corresponding Author:
Darlinton Barbosa Feres Carvalho, BSc, MSc, PhD
Federal University of São João del-Rei
Departament of Computer Science
Prédio de Ciência da Computação, 3° andar
Av Visconde do Rio Preto, s/nº, Colônia do Bengo
São João del-Rei
Brazil
Phone: 55 3379 4935
Email: darlinton@acm.org

Abstract

Background: Alcohol abuse is the primary cause of (public) health problems in most parts of the world. However, it is undeniable that alcohol consumption is a practice that is widely accepted socially in many places, even being protected by law as a cultural and historical heritage. The issue of alcohol abuse is complex and urgent, and consequently, it is necessary to create innovative approaches such as the proposal explored in this study.

Objective: This study aimed to explore the development and evaluation of a serious game for smartphones to present a novel approach to address the issue of alcohol abuse.

Methods: A serious game was developed to instill the consequences of alcohol abuse into the player through experimentation in the game. In the game, the consequences of alcohol use are demonstrated by increasing the game speed that gives an illusion of fun but also leads to a premature death. The evaluation employed an assessment based on the Alcohol Use Disorders Identification Test (AUDIT) and the Game Experience Questionnaire (GEQ). The participants belonged to the university student's house.

Results: The game development process has been presented, including its mechanics and gameplay. The game has the style of action and adventure games in which the player controls an indigenous avatar that can deflect or attack opponents coming his or her way. The game evaluation comprised an assessment based on 23 participants, aged 20 to 29 years. According to the AUDIT assessment, 18 participants reported having a low or nonexistent degree of alcohol dependence and 5 declared average dependence. Regarding their habit of playing games on smartphones, 9 participants declared they have this habit of playing (habitual players), and among the 14 that did not have this habit of playing (nonhabitual players), 3 participants declared not having a smartphone at all. The GEQ core assessment showed a higher positive affect among the participants with a habit of playing games, scoring 2.80 (habitual players) on a scale of 4.0 versus 1.61 (nonhabitual players), and higher tension as an opposite relationship of 0.81 (nonhabitual players) versus 0.37 (habitual players). The overall GEQ evaluation showed that the game presents a more positive than negative affect on all users, besides showing the other desirable characteristics of serious games.

Conclusions: We present a new way of dealing with the issue of alcohol abuse through a game designed for smartphones. It promotes an overall positive user experience, having a greater impact on users accustomed to games. The proposed approach has its niche, though it is still a minority in the evaluated population. Further research should explore new game features, such as new styles, to make the game more attractive to a wider audience, in addition to performing an in-depth study on the effects of playing it.

XSL•FO
RenderX

Introduction

Background

Alcohol abuse is the primary cause of (public) health problems in Brazil. According to the World Health Organization (WHO), in 2012, about 3.3 million deaths or 5.9% of all deaths recorded on the planet were attributable to alcohol consumption [1]. The abusive consumption of alcoholic beverages is the component cause of more than 200 types of diseases and injuries. Alcoholism represents harm to private health, public health, and the economy as a whole. The estimated annual cost of alcoholism is £20 billion in the United Kingdom and more than US $200 billion in the United States. Brazil loses around 7% of its gross domestic product per year because of excessive consumption of alcoholic beverages. The cost estimated in 2014 reached R$372 billion [2].

However, it is undeniable that alcohol consumption is a part of many cultures, being a practice that is widely accepted socially in most parts of the world. In Brazil, the spirit *cachaça* (a type of distilled spirit made from fermented sugarcane, the most popular distilled alcoholic beverage in Brazil) is even protected by law for its distinguished characteristics (Brazilian Presidential decree n.6871/2009) and is considered a part of national cultural and historical heritage. The issue of alcohol abuse is complex and urgent, and consequently, it is necessary to create innovative approaches such as the proposal explored in this study.

Objectives

The main aim of this study was to explore the possibilities offered by new technologies to address this relevant issue. Numerous successful cases have been found in the literature employing serious games in dealing with relevant health care issues [3]. Considering the growing use of smartphones, the creation of a serious game for smartphones takes advantage of this important channel of communication to innovate an intervention for such issues.

Our proposal was to employ new technologies to raise awareness about the consequences of alcohol abuse, promoting empowerment of the individual, traditional cultures, and social responsibility. The authors believe that a suitable means for this is a playful educational tool, that is, a serious game. In this sense, the user is allowed to experience the option of drinking and modifying the gameplay, eventually increasing the fun factor, but also showing the consequences, such as the premature termination of the game (player's death in the game).

Methods

Design Rationale

Game development is based on the premise that games are strongly linked to reality. According to Abt [4], a game has some core elements such as rules, participants, information, gains, and losses. It is also possible to assimilate factors such as competition, opposition, and maximization or minimization of some factor over another. Thus, it is feasible to relate a game to other human activities, as all have rules (of society or a particular environment), participants, successes, and failures (of tasks or procedures involved).

An indigenous hero is considered for the setting of the game. It is worth noting that alcohol abuse also extends in Brazil to the indigenous sphere of society. Langdon [5] points out that because of the process of inclusion of the indigenous people in the broader society, they have begun to substitute or add the distilled beverage into their everyday life. The consequences of this change have also been addressed by Viertler [6] and Guimarães and Grubits [7], demonstrating that alcohol abuse is the main cause of addiction diagnoses, accidents, and cases of violence against other members of the community, including the use of various weapons and incidence of fatal victims. The indigenous theme also brings out the problem in these societies to a wider audience as intervention in this area is also not an easy task and lacks attention in a broader sense.

As Ingold [8] points out, the learning process is about the education of attention. Thus, as a child, learning is consolidated through games and toys. Through them, the study by Ritterfeld et al explains [9], "children exploit and accumulate the world, improving their skills and abilities" because they are extremely motivating activities. However, later in elementary school, there is a clear separation between learning and fun, making learning somewhat unmotivating and unpleasant. This is because motivation dictates the learning flow—when motivation ends, learning and the act of playing also end [10].

Electronic games have been proposed to make learning motivational as they have the ability to communicate concepts and facts of many subjects effectively and allow people to recreate themselves in new worlds and achieve recreation and deep learning at the same time. That is, they are capable of creating a dramatic representation of the studied situation [4,10,11].

Such games are called educational games. However, they do not always achieve their goal as many prioritize the educational or entertainment components separately. A methodology is used specifically for the development of serious games that are games that use the artistic medium of the game to deliver a lesson or teach about some subject so that both components are balanced [9].

According to Michael and Chen [11], a serious game should contain some elements related to design and development itself, including the following: (1) simple (intuitive)—usable by people who have little or no experience with digital games, (2) adequate simulation—both in terms of realism and difficulty, and (3) progress analysis—it allows the player to formulate strategies, in case he or she learns what the game wants to pass to achieve greater progress in the game.

To achieve these desirable characteristics, the development was divided into 3 stages: (1) planning, (2) prototyping, and (3) app consolidation. The whole process was documented in the Game

Development Document, which contains all the necessary information for game development, such as an execution flowchart, the definition of gameplay, and references and inspiration. This paper has presented a consolidated version of the Game Development Document, highlighting the main concepts employed in the game with regard to mechanics, gameplay, and implementation.

Field Evaluation

The game evaluation was carried out with students from the student residence of the Federal University of São João del-Rei. This residence is the student housing provided by the university for low-income students under the university's official assistance program. All participants were volunteers and had to sign a participant consent form, which described the general purpose of the study and stated the research procedures according to international and Brazilian ethical research laws and principles. The results are opinions declared by the participants and were disclosed anonymously. The proposed evaluation consisted of 3 stages: 2 questionnaires (pre- and postintervention) and a 20-min intervention phase of user interaction with the game. No further explanation about the game was provided, leaving the exploration process up to the participants. When the participants asked about a specific game element, the feedback was an incentive for them to explore and test the interface to figure out the answer.

The preintervention interview was performed to characterize the participant by collecting information such as age, sex, and their level of alcohol dependence. To identify the level of alcohol dependence of the users, the Alcohol Use Disorders Identification Test (AUDIT) was employed. The WHO recommends the use of AUDIT [12] as a simple, brief method of screening for excessive drinking. Its questionnaire contains 10 multiple-choice questions with 5 choices each. The assessment works as a sum of points acquired by each alternative—from 0 to 4—and the final result, that is, the sum

of all the alternatives, determines the level of alcohol dependence of the interviewee. The risk degree of dependency, potential consequences, and intervention recommendation for the interviewee are classified according to 4 levels. The correspondence among scores and the meaning of their risk level is shown in Table 1.

The Game Experience Questionnaire (GEQ) [13] was used for the postintervention evaluation. It consists of a questionnaire composed of 3 modules: the core, the postgame, and the social modules. The questionnaire has simple and direct questions about the player's thoughts and feelings, such as "I felt happy" and "I felt angry." The core module has questions regarding feelings and beliefs that the interviewee had while playing. In the postgame module, the questions are focused on when the player stopped playing. In this way, the participant demarcates the level of agreement with the corresponding sentence on a scale from 0 to 4, where 0 means no agreement and 4 means a lot of agreement.

Only the core and postgame modules were used as the proposed game does not have social interactions. The participants answered them after the intervention stage. In each module, a group of questions defined a component to be analyzed, and the final calculation of it was given by the simple average of the questions. The total average of the components was given by the simple average of all the individual components of the interviewees—with results ranging from 0 to 4. In this case, 0 means that there is no presence of that component and 4 means there is a lot of it.

Finally, the postintervention evaluation consisted of 3 questions to evaluate the design of the app concerning the visual and sound resources and the available information about the game (such as tutorials and information about the project) and about the controls. A scale from 0 to 4 was also used at concordance levels, where 0 meant weak and 4 meant optimal.

Table 1. Risk level scores with regard to alcohol consumption behavior according to the Alcohol Use Disorders Identification Test assessment.

Score	Risk level
0-7	Inexistent: low-level drinking or abstinence
8-15	Low: alcohol use in excess of low-risk guidelines
16-19	Medium: harmful and hazardous drinking
Over 20	High: alcohol dependence

Results

Development

The Game Development Document details the proposed modeling and the definition of the rules, basic interface, and other details. The details of the graphics and sound features were being defined along with the game development. In total, 2 versions were developed. The first prototype was made in the Massachusetts Institute of Technology (MIT) AppInventor platform, but its development was discontinued because of performance issues. The last prototype is the evolution of the

first one but using the Unity Graphic Engine (ie, Unity3D) for its implementation.

The game was deployed on the Android operating system because of its popularity. The development of an app through native language is the development of an Android app exclusively using the Android Software Development Kit or Android Native Development Kit. This allows for efficient development as the developer has direct access to all the tools of the device. Access to all tools allows you to design the app in the best possible way for the operating system, maximizing the user experience. However, this environment requires mastery of the platform's native language and its tools. All physical modeling involved in the app must be developed. For easy game

XSL·FO

RenderX

development, many platforms can be found in the Android ecosystem that assist developers in creating new games.

A study was carried out on game development platforms available for the Android system. The most promising ones were MIT AppInventor and Unity Graphic Engine.

The MIT AppInventor platform is an Android app development platform supported by Google and maintained by the MIT (United States). It has a drag-and-drop system where users select, drag, and fit blocks that determine a line or method of execution in an interaction area. It is an open-source license environment that is easy to learn, with the freedom to run the program for any legal purpose.

At first, the authors chose the MIT AppInventor platform for prototyping because of the small learning curve and the need to set implementation details. It was a straightforward implementation of what was designed. However, because of performance issues when the game had to perform fast updates of the graphic elements on the screen (ie, the game starts experiencing unexpected screen update errors), a more robust platform was needed. Therefore, the game's final version was fully implemented using the Unity Graphic Engine. It took a while to learn this new platform, but the process of prototyping in AppInventor to achieve better implementation later in Unity was shown to be effective to enable fast prototyping, especially considering programmers without much experience in game development, and produce a better game in terms of quality on a more robust platform.

The proposed game follows the style of action and adventure games, in which the player controls an indigenous avatar who can deflect or attack opponents coming his or her way. It has 3 screens: the initial menu screen, an information display screen about game rules and the project, and the screen that contains the game. The initial menu screen, described in Figure 1 (left), allows you to move to any other screen within the app and to exit the app. The about screen, shown in Figure 1 (right), allows the user to return to the initial menu screen (using the return button in the upper left corner). Information regarding the game screen (and the game itself) is described in the following subsections. It is important to note that all graphics and sounds are available on the internet under the Creative Commons license.

The instructions presented in the about screen (Figure 1 [right]) can be translated as follows: *About the Game* – "JIB is a game created to address the issue of alcoholism, allowing the player to experience the consequences of alcohol abuse through the game. The indigenous theme has been explored to compose the hero of this game, creating a narrative contextualized in the forest. The effects of using alcohol can be perceived by increasing the game speed and giving an illusion of more fun but also by showing that alcohol abuse might lead to premature death."; and *How to Play* "Move the Indian by touching the lower half left and right of the screen and shoot by touching the center of the screen. Opponents will appear at the top of the screen, and you must act. Be careful with the drinks, it can be tasty, but the alcohol level will rise, and it will change you."

Figure 1. Two screens of the proposed game: (left) first screen; (right) about the game screen (in Portuguese).


The user interaction consists of using a touch screen, where areas are defined for interaction and arrangement of the elements of the game. Figure 2 (right) presents a schematic of these areas in comparison with the game screen (Figure 2 [left]: (1) move character to the left, (2) shoot the arrow, (3) move character to the right, (4) restart the game, (5) enable or disable the sound in the game, (6) enable or disable the display of indicative signs of the action buttons, (7) display the player's current score, (8) return to the home screen, and (9) indicate current alcohol consumption level).

The player, as already mentioned, controls an indigenous avatar who can deflect or shoot arrows at the opponents that appear in his or her way. Opponents appear on the upper edge of the device screen, and when they reach the lower end, they disappear. Some defined opponents are the snake, the jaguar, and the drink. Once the Indian comes in direct contact with the snake or the jaguar, the game ends. If the arrow hits an opponent, the player gets a score that is added and displayed on the screen.

When the player consumes the drink, it increases the points and speed of the game, which in a way also increases the fun. The idea is that drinking causes the player to slow down compared with the world, that is, the other elements get faster.

Figure 2. User interface with interaction elements positioning: (left) game play screen; (right) sensible areas of the interface for user interaction.
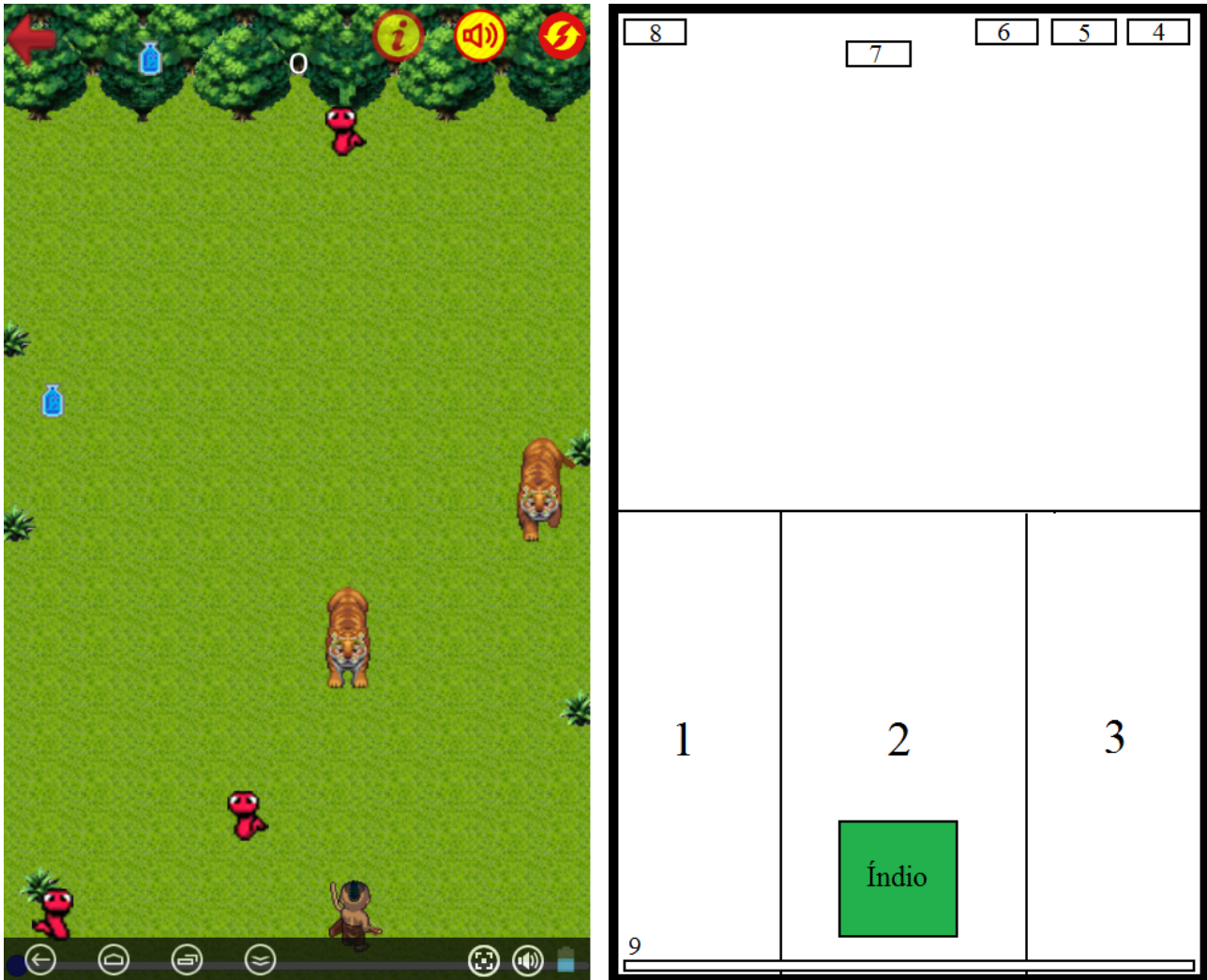
The opponents use a system for continuous displacement, explained in Figure 3, to avoid the overlap and improper collision of opponents on the screen. The lanes have counters associated with them to notify their availability (or not). At the launch of an opponent, the counter of the chosen and adjacent lane is incremented. Counters are decreased when an opponent passes a safety range (represented by the horizontal blue dashed line closer to the top in Figure 3).

Figure 3. Enemy movement lanes in the game screen.
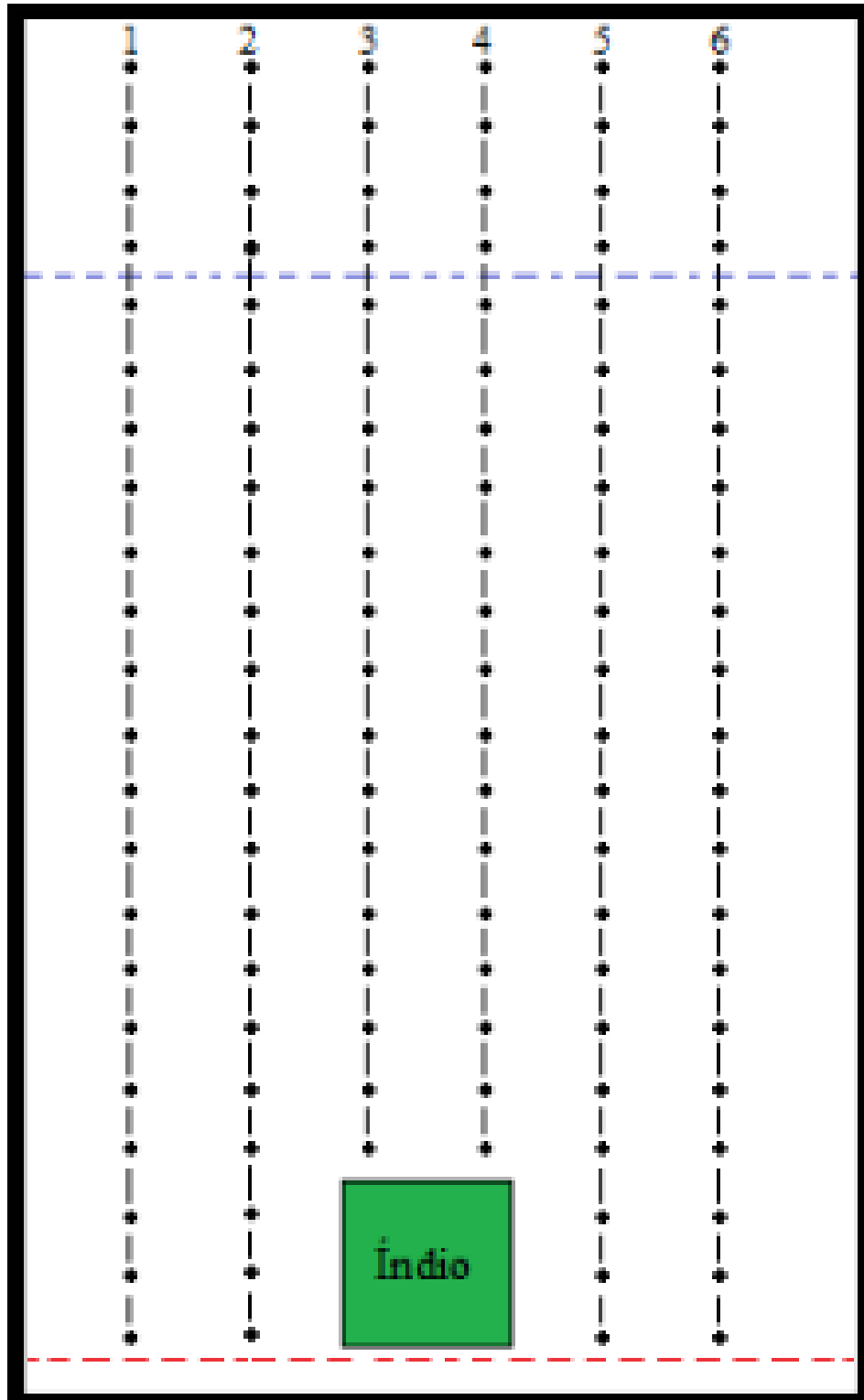
Evaluation

Data were collected from a total of 23 participants, 12 of whom were male and 11 were female. The participants' ages ranged from 20 to 29 years, although 2 participants preferred not to declare their age. Regarding the degree of dependency identified by the AUDIT questionnaire [12], 5 claimed not having any alcoholic dependence, 13 reported having a low or nonexistent degree of dependence, and 5 declared an average dependence. No participant declared a high degree of dependency.

Of the 23 participants, only 3 did not have a smartphone. Among those who owned a smartphone, 11 did not use it for games.

Game Experience Questionnaire

The GEQ [13] was used to gain an understanding of user experience with the game. Overall, 2 GEQ modules were used: core and postgame.

First, the elements of the questionnaire should be understood. According to Johnson et al [14], the components—competence, stress, negative affect, and positive affect—are self-explanatory. Negative affect and positive affect refer to the user experience. For example, a user may feel good or unmotivated about the game. Challenge aims to present data regarding the amount of effort, difficulty, and pressure felt by the user during the game. Tension is related to the player's frustration while playing the game. Flow aims to identify how interested the user was during the game. Immersion shows how immersed the user was with the story and elements of the game. Finally, competence is related to the player's ability and how well he or she performed during the game. All core elements of the GEQ are presented in Table 2.

The GEQ postgame questionnaire was used to identify how the player felt after playing. This questionnaire is composed of 4 variables: positive affect, negative affect, tiredness, and return to reality. Positive affect is related to the satisfaction, victory, and power of the user after playing. In contrast, the negative affect addresses the user's bad experiences after playing. Tiredness is related to the user's exhaustion during the game. Finally, the returning to reality component addresses the user's disorientation after a gaming period. Table 3 presents the postgame elements and related questions.

Table 2. Core Game Experience Questionnaire components and response options.

Core Game Experience Questionnaire components	Response options
Immersion	- I was interested in the game's story - It was aesthetically pleasing - I felt imaginative - I felt that I could explore things - I found it impressive - It felt like a rich experience
Flow	- I was fully occupied with the game - I forgot everything around me - I lost track of time - I was deeply concentrated on the game - I lost connection with the outside world
Competence	- I felt skillful - I felt competent - I was good at it - I felt successful - I enjoyed it
Negative affect	- It gave me a bad mood - I thought about other things - I found it tiresome - I felt bored
Tesnion	- I felt annoyed - I felt irritable - I felt frustrated
Challenge	- I thought it was hard - I felt pressured - I felt challenged - I felt time pressure - I had to put a lot of effort into it

Table 3. Postgame Game Experience Questionnaire components and response options.

Postgame Game Experience Questionnaire components	Response options
Positive affect	• I felt revived • It felt like a victory • I felt energized • I felt satisfied • I felt powerful • I felt proud
Negative affect	• I felt bad • I felt guilty • I found it a waste of time • I felt that I could have done more useful things • I felt regret • I felt ashamed
Tiredness	• I felt exhausted • I felt weary
Returning to reality	• I found it hard to get back to reality • I felt disoriented • I had a sense that I had returned from a journey

Analysis

Analyses of the results from the core and postgame GEQ questionnaire (Tables 2 and 3, respectively) are presented as follows. The mean score was calculated for each component in both questionnaires. The average score was calculated for 3 different classes of users: nonhabitual players, habitual players, and total. Table 4 presents the results of the core questionnaire for the 3 distinct classes.

As shown in Table 4, the highest score found among users who declared themselves to be nongamers was the positive affect component (nonhabitual players=1.61), followed by flow and competence components (nonhabitual players=1.54). However, although the positive affect component obtained a higher score, it was similar to the negative affect component (nonhabitual players=1.25). Tension component was the one with the lowest score (nonhabitual players=0.81). The scores of the components, competence (nonhabitual players=1.54), immersion (nonhabitual players=1.44), and challenge (nonhabitual players=1.33), were higher than the general mean (mean 1.25).

The highest score was of the positive affect component (habitual players=2.8), followed by competence (habitual players=2.18), immersion (habitual players=1.96), flow (habitual players=1.89), challenge (habitual players=1.16), negative affect (habitual players=0.69) and, finally, tension (habitual players=0.37). It should be noted that in this evaluated class, the positive affect component scored considerably higher than the negative affect component. The scores obtained in the components, tension, negative affect, and challenge, were below the arithmetic mean (mean 1.57).

A combination of the 2 classes (nonhabitual players and habitual players) was also performed. The positive affect component had the highest score (total=2.08), followed by the components, competence (total=1.79), flow (total=1.68), immersion

(total=1.64), challenge (total=1.26), negative affect (total=1.03), and, finally, tension (total=0.64). In the combination of the 2 classes, it can be noted that the positive affect component had a relatively higher score than the negative affect component. Only the components, challenge, negative affect, and tension, reached scores below the mean (mean 1.44).

With regard to the results obtained through the GEQ postgame questionnaire for nonhabitual players, the positive affect component reached the highest score (nonhabitual players=0.75). Negative affect was lower than the positive affect (nonhabitual players=0.5). The lowest value found was related to the return to reality component (nonhabitual players=0.43). The tiredness component obtained a relatively high value in comparison with the other components (nonhabitual players=0.68). The results referring to the GEQ postgame questionnaire for the 3 distinct classes are presented in Table 5.

For the habitual players class, the component with the highest score was the positive affect (habitual players=1.17), followed by the components return to reality (habitual players=0.48), negative affect (habitual players=0.2) and, finally, tiredness (habitual players=0). Positive affect reached a score considerably higher than negative affect. The components below the arithmetic mean (mean 0.46) were negative affect and tiredness.

The following scores were obtained for the 2 combined classes (nonhabitual players and habitual players): positive affect (total=0.95), return to reality (total=0.47), tiredness (total=0.43), and negative affect (total=0.40). Only the positive affect component had a score above the arithmetic mean (mean 0.56). However, this score was responsible for an increase in the mean.

The users also evaluated the proposed game interface in 3 aspects: (1) defined controls for the game, (2) quality of the graphics and sound, and (3) game information about the project and how to play. The simple arithmetic mean of these aspects was calculated and is presented in Table 6.

Table 4. Core Game Experience Questionnaire average scores for the 3 classes.

Core Game Experience Questionnaire	Nonhabitual players, mean	Habitual players, mean	Total, mean
Negative affect	1.25	0.69	1.03
Positive affect	1.61	2.80	2.08
Challenge	1.33	1.16	1.26
Tension	0.81	0.37	0.64
Flow	1.54	1.89	1.68
Immersion	1.44	1.96	1.64
Competence	1.54	2.18	1.79

Table 5. Postgame Game Experience Questionnaire average scores for the 3 classes.

Core Game Experience Questionnaire	Nonhabitual players, mean	Habitual players, mean	Total, mean
Positive affect	0.75	1.17	0.95
Negative affect	0.50	0.20	0.40
Returning to reality	0.68	0.68	0.43
Tiredness	0.43	0.48	0.47

Table 6. The average of usability aspects based on evaluation.

Aspect	Values (mean)
Game control	2.86
Graphics and audio resources	3.09
Available information	2.05

An important feature of this experiment was the division of participants into those who are habitual players and those who are not (nonhabitual players). It should be taken into account that the analyses performed by *players* tend to produce less variation as these participants have a basis for comparison with previous experiences.

During analysis of the results, it was verified that the participants more accustomed to playing games did not feel challenged by the Jib game. This can be explained by the difficulty of the game as the game does not have an increasing level of difficulty and can be saturated very quickly, using few iterations. Another factor that potentially influenced the perceived level of challenge in the game was self-declared competence.

The score of the immersion component was significant, even though the game was simple. Participants used to play games reported a good experience in this component. The returning to reality component scored 0 among participants who are habitual players. The authors believe that the habit of playing more complex games had a great influence on this component as the sense of reality of these people may be less influenced by Jib, and they could easily discern their virtual experience from the real one (returning to reality). In such cases, serious games with low complexity can possibly generate more significant results in people who are nonhabitual players.

In total, 2 participants were identified as outliers (participant 11 and 22). Both of them had much higher stress levels (tension) than the other participants. This component possibly had an

influence on the negative and positive affect components. Participant 22 had the highest negative affect score among all the participants. Participant 11, on the contrary, obtained a score for positive affect within the average, which was not expected, given his or her level of tension.

Discussion

Related Studies

Although there is a vast body of literature available regarding serious games applied to health care, this is a recent topic, and therefore, there are only a few pieces in the literature related to serious games that deal with alcohol abuse.

Gaibler et al [15] presented a study that addresses the effect of alcohol on safe driving activity, especially in the young age group (between 14 and 27years). The app, a serious game, is a racing game in the third person, where the intention is to reach the final goal safely, avoiding alcoholic beverages in the process. Our proposal shares the principle of alcohol consumption affecting the gameplay. In this game, the players are challenged by lowering their game vision according to the level of intoxication. In our proposal, varying the game speed offers a more dramatic consequence to the user experience, leading to a premature end of the game. Nonetheless, this paper is limited to a discussion of some preliminary and promising results.

Rodriguez et al [16] conducted a systematic review related to serious educational games aimed at the consumption of alcohol and other drugs by adolescents. The search for papers was

XSL·FO

RenderX

carried out in research portals. According to the authors, 8 papers related to the consumption of alcohol, cannabis, tobacco, methamphetamine, ecstasy, and other drugs were found. In addition, 6 other papers addressed the use of other drugs. However, only one of the papers showed a decrease in the frequency of drug use. The authors highlight the need for further investigation and development of serious educational games, such as the one presented in this paper.

Boendermaker et al [17] applied gamification techniques similar to a cognitive bias modification of attention (CBM-A) training task to draw attention away from images of alcoholic beverages. The applied training task is called a visual probe task [18,19] that consists of the use of pairs of images, where one presents an important stimulus to the alcoholic beverage and another one a neutral stimulus (something nonalcoholic). The activity consisted of 4 sessions, with at least a 1-day difference between sessions, for 2 weeks. The study was aimed at undergraduate students (96 students, mean age 21.2 years). The authors identified the problems of excessive alcohol consumption of the candidates by using questionnaires. The authors concluded that the innovation proposed in their study was insufficient to the task of motivating adolescents in training when compared with conventional CBM-A training. They believe that one of the motivations for such an outcome is related to the expectation that a game should be fun, a feature that was not the focus of their study. In this study, the focus is on building a game that is fun and fosters educational drive, but no analysis regarding user motivation to play the game or gameplay consequences has been performed.

More recently, Boendermaker et al [20] developed and evaluated a serious game aimed at increasing behavioral control in adolescents and thereby helping them to improve control over their alcohol use. The studied compared the game training to a game placebo and a nongame training version in a randomized controlled trial. The study sample was 185 adolescents (mean age 14.9 years) assessed for 4 weeks. As a result, the game variants were shown to motivate adolescents beyond the level of the nongame version. The exercised skill, behavioral control, improved significantly over time, but this effect was also present in the game placebo. Yet, as baseline drinking levels were low, no significant training effects on drinking behavior were found. Our study also employs a sample that has been characterized by having a low baseline on drinking. The authors believe the game activities alone may have had a beneficial impact on the measures of behavioral control.

Although the current results are not yet conclusive as to whether their game is effective regarding its purpose, the study shows exciting findings of the suitability of serious games. Serious games seem to be suitable to exercise new skills, but the evaluation of its effectiveness is still a major challenge. Nevertheless, their results, and our results, suggest that this is a promising way, and further research is required to disclose and assess its potential.

Conclusions

The paper presented the development of a serious game for smartphones aimed at addressing the issue of alcohol abuse. The main purpose was to call the player's attention to the consequences of alcohol abuse through experimentation in the game. The authors do not have the illusion that people would cease drinking immediately but do intend to contribute to the awareness of this social problem.

To evaluate the proposed approach, an assessment based on AUDIT and GEQ was carried out with residents of a student's house. The proposed evaluation consisted of 3 stages: 2 questionnaires (pre- and postintervention) and 1 intervention stage, which consisted of 20 min of exposure to the game. The users were divided into groups based on their habit of gaming in the data analysis.

The quantitative analysis presents a high degree of positive affect concerning the participants who declared themselves to have the habit of playing, although they did not have a high level of challenge because of the low difficulty and small learning curve. Overall (among all participants), the positive affect score was twice that of the negative affect. In this way, the authors believe that the Jib game obtained a proper evaluation in the 2 different groups of participants, considering the limitations of this first version of the game.

In the future, the authors expect to continue to develop Jib and to increase the level of difficulty and application of the game. New experiments should also be conducted using different groups of users, such as indigenous people in urban contexts, and exploring the variations between gender, age, and family background related to the issue of alcohol abuse. The net effect of the gaming experience, with the implied alcohol use and fun, needs to be better studied and its suitability verified especially regarding addicts. Further research should also explore new game features, such as new styles, to make it more attractive to a wider audience, in addition to performing an in-depth study on the effects of playing the game.

Conflicts of Interest

None declared.

References

XSL·FO

RenderX

1. World Health Organization. Global Status Report on Alcohol and Health. Geneva, Switzerland: World Health Organization; 2014.

2. Correio Braziliense. 2015. Brazil Loses 7.3% of PIB With Excessive Alcohol Consumption URL: http://www.correiobraziliense.com.br/app/noticia/economia/2015/01/18/internas_economia,466870/brasil-perde-7-3-do-pib-com-consumo-excessivo-de-bebidas-alcoolicas.shtml [accessed 2018-02-05] [WebCite Cache ID 6x06jpO9Y]

3. Arnab S. In: Dunwell I, Debattista K, editors. Serious Games For Healthcare: Applications And Implications. Hershey, Pennsylvania, USA: IGI Global; 2012.

4. Abt CC. Serious Games. Lanham, Maryland: University Press Of America; 1987.

5. Langdon EJ. What to Drink, How to Drink, and When to Drink: The Sociocultural Context of Alcoholism Among Indigenous Populations. In: Proceedings of the Seminar on Alcoholism and STD / AIDS Vulnerability among the Southern, Southeast, and Mato Grosso do Sul People. 2001 Presented at: Seminar on Alcoholism and STD / AIDS Vulnerability; 1997; Londrina, PR, Brazil p. 83-97 URL: http://bvsms.saude.gov.br/bvs/publicacoes/021anais_seminario.pdf

6. Viertler RB. Interethnic living and alcoholism among the Bororo: results of a survey. Rev Tellus 2002 Apr;2(2):9-38 [FREE Full text] [doi: 10.20435/tellus.v0i2.8]

7. Guimarães LA, Grubits S. Alcoholism and violence in indigenous ethnic groups: a critical view of the Brazilian situation. Psicol Soc 2007;19(1):45-51. [doi: 10.1590/S0102-71822007000100007]

8. Ingold T. From the transmission of representations to the education of attention. In: Whitehouse H, editor. The Debated Mind: Evolutionary Psychology Versus Ethnography. Oxford, United Kingdom: Berg Publishers; 2001:113-153.

9. Cody M, Vorderer P, Ritterfeld U. Serious Games: Mechanisms and Effects. New York: Routledge; 2009.

10. Gee JP. What Video Games Have to Teach Us About Learning and Literacy. Basingstoke, United Kingdom: Palgrave Macmillan; 2003.

11. Michael D, Chen S. Serious Games: Games That Educate, Train, and Inform. Boston, MA, USA: Course Technology PTR; 2006.

12. Babor F, Higgins-Biddle JC, Saunders JB, Monteiro MG. World Health Organization. 2001. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care URL: https://www.who.int/substance_abuse/publications/audit/en/ [accessed 2018-02-01] [WebCite Cache ID 6wuXNWsa1]

13. Ijsselsteijn WA, de Kort YA, Poels K. Eindhoven University of Technology. 2013. The Game Experience Questionnaire URL: https://research.tue.nl/en/publications/the-game-experience-questionnaire [accessed 2019-09-11]

14. Johnson D, Wyeth P, Sweetser P, Gardner J. Personality, Genre and Videogame Play Experience. In: Proceedings of the 4th International Conference on Fun and Games. 2012 Presented at: FnG'12; September 4-6, 2012; Toulouse, France p. 117-120. [doi: 10.1145/2367616.2367633]

15. Gaibler F, Faber S, Edenhofer S, von Mammen S. Drink & Drive: A Serious but Fun Game on Alcohol-Induced Impairments in Road Traffic. In: 7th International Conference on Games and Virtual Worlds for Serious Applications. 2015 Presented at: VS-Games'15; September 16-18, 2015; Skovde, Sweden p. 1-5. [doi: 10.1109/VS-GAMES.2015.7295771]

16. Rodriguez DM, Teesson M, Newton NC. A systematic review of computerised serious educational games about alcohol and other drugs for adolescents. Drug Alcohol Rev 2014 Mar;33(2):129-135. [doi: 10.1111/dar.12102] [Medline: 24329810]

17. Boendermaker WJ, Maceiras SS, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. J Med Internet Res Serious Games 2016 Dec 6;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]

18. MacLeod C, Rutherford E, Campbell L, Ebsworthy G, Holker L. Selective attention and emotional vulnerability: assessing the causal basis of their association through the experimental manipulation of attentional bias. J Abnorm Psychol 2002 Feb;111(1):107-123. [doi: 10.1037/0021-843X.111.1.107] [Medline: 11866165]

19. Schoenmakers TM, de Bruin M, Lux IF, Goertz AG, van Kerkhof DH, Wiers RW. Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. Drug Alcohol Depend 2010 Jun 1;109(1-3):30-36. [doi: 10.1016/j.drugalcdep.2009.11.022] [Medline: 20064698]

20. Boendermaker WJ, Veltkamp RC, Peeters M. Training behavioral control in adolescents using a serious game. Games Health J 2017 Dec;6(6):351-357. [doi: 10.1089/g4h.2017.0071] [Medline: 28926286]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
CBM-A: cognitive bias modification of attention
GEQ: Game Experience Questionnaire
WHO: World Health Organization

XSL•FO
RenderX

Original Paper

Effect of Brief Biofeedback via a Smartphone App on Stress Recovery: Randomized Experimental Study

John F Hunter[1], PhD; Meryl S Olah[1], BA; Allison L Williams[2], PhD; Acacia C Parks[2], PhD; Sarah D Pressman[1], PhD

[1]Department of Psychological Science, University of California, Irvine, Irvine, CA, United States
[2]Happify Health, New York, NY, United States

Corresponding Author:
John F Hunter, PhD
Department of Psychological Science
University of California, Irvine
4201 Social & Behavioral Sciences Gateway
Irvine, CA, 92697
United States
Phone: 1 9493459987
Email: jhunter1@uci.edu

Abstract

Background: Smartphones are often vilified for negatively influencing well-being and contributing to stress. However, these devices may, in fact, be useful in times of stress and, in particular, aid in stress recovery. Mobile apps that deliver evidence-based techniques for stress reduction, such as heart rate variability biofeedback (HRVB) training, hold promise as convenient, accessible, and effective stress-reducing tools. Numerous mobile health apps that may potentially aid in stress recovery are available, but very few have demonstrated that they can influence health-related physiological stress parameters (eg, salivary biomarkers of stress). The ability to recover swiftly from stress and reduce physiological arousal is particularly important for long-term health, and thus, it is imperative that evidence is provided to demonstrate the effectiveness of stress-reducing mobile health apps in this context.

Objective: The purpose of this research was to investigate the physiological and psychological effects of using a smartphone app for HRVB training following a stressful experience. The efficacy of the gamified Breather component of the Happify mobile health app was examined in an experimental setting.

Methods: In this study, participants (N=140) underwent a laboratory stressor and were randomly assigned to recover in one of three ways: with no phone present, with a phone present, with the HRBV game. Those in the *no phone* condition had no access to their phone. Those in the *phone present* condition had their phone but did not use it. Those in the *HRVB game* condition used the serious game Breather on the Happify app. Stress recovery was assessed via repeated measures of salivary alpha amylase, cortisol, and self-reported acute stress (on a 1-100 scale).

Results: Participants in the *HRVB game* condition had significantly lower levels of salivary alpha amylase during recovery than participants in the other conditions ($F_{2,133}=3.78$, $P=.03$). There were no significant differences among the conditions during recovery for salivary cortisol levels or self-reported stress.

Conclusions: These results show that engaging in a brief HRVB training session on a smartphone reduces levels of salivary alpha amylase following a stressful experience, providing preliminary evidence for the effectiveness of Breather in improving physiological stress recovery. Given the known ties between stress recovery and future well-being, this study provides a possible mechanism by which gamified biofeedback apps may lead to better health.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

Although smartphones are often criticized for contributing to ill-being [1-4], these devices hold great potential for improving one's well-being if utilized properly in specific contexts. Smartphones may be particularly useful as tools that provide gamified apps to deliver stress-buffering interventions. Stress is prevalent in many peoples' lives, and its accumulated effects can lead to various undesirable physical and mental health outcomes, such as an increased risk for mortality [5]. Many of these negative outcomes are due to prolonged activation of one's stress systems [6]. However, if individuals employ strategies that promote more efficient recovery from stressors, some of the negative long-term impacts may be mitigated.

Why do smartphones present a promising opportunity for altering stress recovery? Psychologists have developed a variety of evidence-backed strategies that aid in stress reduction [7], such as biofeedback training, which directs individuals to monitor and attempt to alter their physiological arousal pattern [8]. Smartphones are ideally situated to be used as a tool for biofeedback training and to combat the negative effects of stress because they are popular, conveniently accessible, and have an array of technological capabilities [9]. Since these devices are nearly omnipresent in daily life, they can deliver interventions and assistance wherever and whenever needed.

Smartphones provide a range of possibilities for helping individuals recover from a stressful experience and may even do so when not actively used. Even when merely present, smartphones serve as symbols that can cause cognitive distraction [10] or activate representations of social connections [11]. Distraction induced by smartphones has generally been viewed as detrimental [12]; however, such distraction can be beneficial when faced with a stressor, because it can draw attention away from the negative stimuli at hand and help circumvent rumination [13]. In addition, symbolic representations of social connections can elicit perceptions of social support [14], which, when perceived passively, is the most effective form of support for stress alleviation [15]. By providing distraction and perceived social support, the mere presence of a smartphone may aid in stress recovery by serving as a "digital security blanket" in instances of social stress [16]. Thus, it is important to investigate more fully how merely having a smartphone in one's presence may aid in stress alleviation.

Research has demonstrated that *actually using* one's smartphone can be beneficial or detrimental for stress recovery, depending on how and when the device is used. For example, research has shown that using social media sites such as Facebook can provide social resources that sometimes help buffer acute stress [17], but at other times, fail to do so [18]. In some instances, social support gleaned via text message can reduce cardiovascular responses to stress [19]. However, sending and receiving text messages can also increase physiological indicators of stress such as heart rate, respiration, and skin conductance [20]. These mixed findings concerning how phone use influences stress underscore the fact that the ways in which we commonly interact with our devices are not universally

beneficial for stress recovery. In fact, when looked at more broadly, greater use of smartphones is associated with higher levels of physiological stress [21]. Therefore, if we hope to highlight the most effective ways to use a phone to reduce stress, it may be important to go beyond natural phone use habits and, instead, provide structured apps that are specifically designed for stress reduction.

One promising way to use a smartphone to aid in stress reduction is by engaging with a mobile health (mHealth) app. mHealth apps can utilize technological capabilities (eg, phone sensors, interactive displays) and draw on the ubiquity of smartphones in everyday life to deliver functional and convenient interventions [22]. By combining evidence-based stress-reduction techniques with an engaging and ever-present medium, mHealth apps hold great promise for mitigating the negative effects of stress.

Happify is an mHealth app that provides gamified activities aimed at improving well-being and reducing stress [23]. Happify is representative of multiple aspects of other mHealth apps because it employs various smartphone technological capabilities (eg, sensors, visual and audio components, engaging interface) and incorporates empirically validated strategies to deliver training in a self-contained package. Within the Happify suite of activities, the Breather function delivers heart rate variability biofeedback (HRVB) training (Multimedia Appendix 1). HRVB is a particularly effective stress-reducing activity that targets changes in heart rate variability (HRV) by regulating breathing and bringing awareness to physiological function [24]. HRV is an index of beat-to-beat changes in heart rate and is an indicator of parasympathetic nervous system activity [25]. When undergoing a stressor, the typical response is for our sympathetic nervous system to activate and parasympathetic activity to decline (indicating low HRV). However, an adaptive response to a stressor would be for an individual to exhibit higher HRV. This is because greater fluctuations in heart rhythm (higher HRV) indicate greater adaptability to physiological needs than fewer fluctuations (lower HRV) [24]. When HRV is higher, it is a sign that our cardiovascular system (and multiple associated systems) is responding appropriately to environmental demands (eg, a stressor). Thus, using HRVB to increase HRV may be helpful when recovering from a stressor because it activates our parasympathetic nervous system and allows us to more quickly reduce physiological arousal. It is also important to note that high HRV is considered a protective factor against cardiovascular disease and is generally associated with good health and well-being [26]. Additionally, low HRV has numerous negative implications for long-term health outcomes, such as increased risk for mortality and morbidity [27,28]. HRVB training has been successfully utilized in a variety of acute stress settings and is well-validated technique for reducing stress [29,30]. The goal of undergoing a 5-minute guided session on Breather is for the user to increase HRV and recover effectively from a stressful experience.

Delivering HRVB through a smartphone app provides many advantages over traditional training. Breather overcomes barriers of nondigital HRVB interventions (eg, bulky and expensive equipment, lengthy sessions) because it is quick to administer, is portable and readily accessible, and has all the hardware and

software integrated into a single device. Breather has taken advantage of mobile technology affordances to package an HRVB product in ways that should allow it to be used across a variety of stressful contexts.

Users of Breather generate HRV observations by placing their index finger over the camera of their smartphone (Figure 1A). The light from the camera can be used to monitor blood volume changes within the finger. This process (ie, photoplethysmography) relies on measuring changes in light absorption on the skin of the finger. Algorithms programmed by Happify software engineers then transform those data into a simple signal that is visible to the user. The accuracy of this technology for determining HRV has been recently validated in a series of experiments that compared simultaneously obtained HRV metrics from Happify Breather and traditional electrocardiogram techniques using electrodes [31].

After calibrating the heart rate of the individual, a circular meter directs the individual to follow the breathing patterns on screen (Figure 1B). The meter directs the individual to breathe in for 4 seconds and then breathe out for 6 seconds. This 10-second breathing cycle is ideal for creating a resonant frequency (ie, breathing and heart rate align) that should maximize HRV [32].

After calibration is complete, the interface changes into a calming nature scene (eg, underwater coral bed, tropical beach, mountaintops). The user then travels through the natural environment while he/she continues to breathe along with the meter (Figure 2A). As they breathe deeply and regularly, their HRV increases and the scene becomes more complex and beautiful (eg, coral polyps bloom, flowers grow; Figure 2B). By visually monitoring the changes in the scene, individuals are undergoing HRVB; this process is analogous to how individuals monitor electrocardiogram signals in more traditional "nongamified" HRVB trainings. The app is designed to increase HRV and reduce stress if the users adhere to the directions properly for a 5-minute session.

Figure 1. On-screen instructions for obtaining heart rate variability measurement and calibrating breathing guidance.
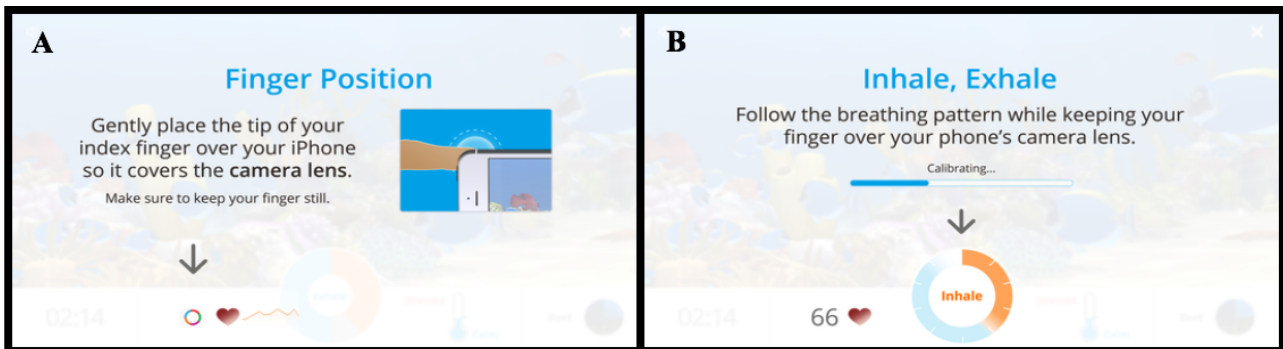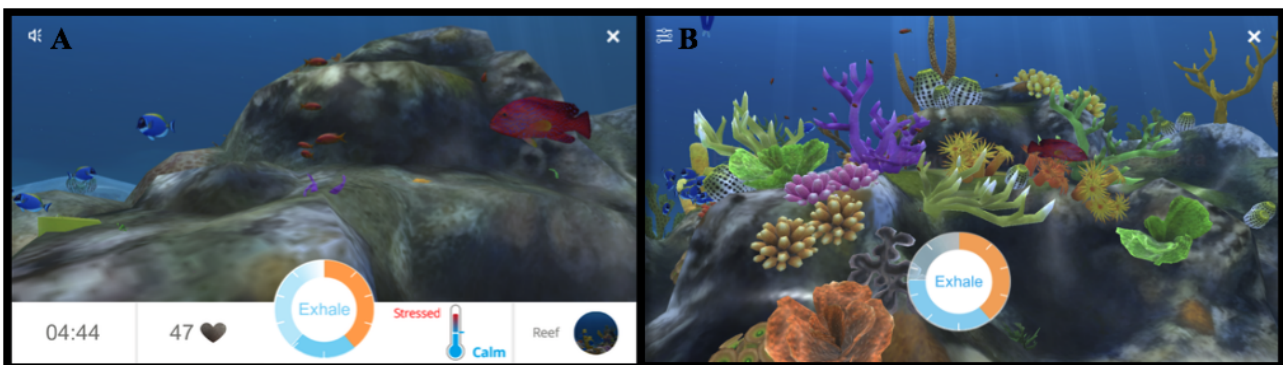

Figure 2. Example of how the full display unfolds when an individual is using Breather.

Assessment of Stress

In this paper, we used a multimethodological approach to determine the effects of smartphones on stress recovery. Stress can be assessed in a variety of ways, and each method provides unique insight into the complex dynamics of how stress impacts our bodies and brains. One of the most common ways to assess stress is to ask individuals to subjectively rate their stress level. Although self-report is advantageous for assessing perceived stress, there are problems of bias (eg, self-presentation concerns) that limit the accuracy and generalizability of these assessments [33]. Due to its complexity, the most appropriate and comprehensive manner in which stress is assessed is a

multimodal approach that combines subjective and objective assessments [34].

One of the most effective, reliable, and efficient ways to capture physiological measurements of stress is to analyze salivary biomarkers. Salivary cortisol is a downstream output of hypothalamic-pituitary-adrenal axis system activation and is one of the most widely used and reliable measures of physiological stress [35]. Higher levels of salivary cortisol indicate greater physiological stress. Another emerging indicator of physiological stress is salivary alpha amylase (sAA), which is an indicator of autonomic nervous system activity and is most strongly tied with sympathetic nervous system activity, the system responsible for the "fight-or-flight" response [36]. When

physiologically aroused, sAA is released via the salivary glands and indicates an immediate stress response. In some cases, sAA is more strongly tied to stress and anxiety than cortisol [37,38], and sAA (but not cortisol) has also been shown to be influenced by smartphones while recovering from a stressor [16]. Thus, sAA is considered our primary outcome of interest. These salivary assessments of hypothalamic-pituitary-adrenal and autonomic nervous system activities combined with self-report give researchers a comprehensive understanding of physiological responses to stress.

This Study

This study investigated the effectiveness of using an HRVB smartphone app to aid in stress recovery. In order to account for the potential stress-buffering effects of simply having a phone [16], we included a condition in which individuals had a phone in their presence. Thus, use of the HRVB app was compared to two control conditions, one in which no phone was present and one in which individuals had their smartphones present when recovering from a stressor.

To examine how these different types of smartphone interactions influence stress recovery, a laboratory experiment was conducted in which participants underwent a standardized stressor, used their phone in a particular way depending on their assigned condition, and were assessed on a range of psychological and physiological stress indicators. We hypothesized that those in the *HRVB game* condition would recover from the stressor more effectively than those who had their phones present or had no phone at all.

This study is one of the first empirical investigations to assess the effects of smartphone app usage on salivary biomarkers of physiological stress [39]. To our knowledge, it is also the first to examine the stress-buffering effects of an HRVB intervention delivered via a smartphone without any external equipment. In addition, since simply having a phone in your presence has been shown to aid in stress recovery, the inclusion of separate experimental conditions for a HRVB game and mere phone presence enabled us to differentiate their effects on stress. The results of this study will help our understanding of why smartphones might be helpful in times of stress, which may inform future recommendations about the most effective way to use a smartphone following a stressful experience.

Methods

Participants

The study was approved by the University of California, Irvine Institutional Review Board, and participants were recruited via the University of California, Irvine undergraduate psychology subject pool. These data were drawn from a larger project that included additional research questions outside the scope of this study. For this particular study, a total of 140 participants were examined (mean age 20.28, SD 2.68; 77.1% female; 45.7% Asian; 27.9% Hispanic/Latino; 15.7% Caucasian; 6.4% African American). Participants were screened for eligibility and excluded from participation if they were diagnosed with a cardiovascular disease, were regularly taking mood altering or cardiovascular altering medication, regularly smoked cigarettes,

were not fluent in English, or did not have an iPhone. All participants were University of California, Irvine, students and consented to participate. Data collection took place from July 2018 through February 2019.

Procedures

Participants underwent an approximately 90-minute laboratory session. All participant phones were confiscated at the beginning of the study under the pretext of measuring the external physical properties of the phone, which allowed the experimenter to later manipulate the phone conditions without arousing suspicion and ensure that all participants experienced similar circumstances of having their phone taken away. For participants randomly assigned to the *HRVB game* condition, the Happify app was installed on their phone and the experimenter guided them through the calibration settings of the Breather function while carefully concealing any indication that the purpose of using Breather was to reduce stress. Participants in the other conditions filled out surveys during this time. After participants completed a series of questionnaires and acclimated to the laboratory environment (approximately 25 minutes), the experimenter returned to the laboratory room and collected a baseline saliva sample. Participants were instructed in the passive drool technique of collecting their own saliva sample.

Participants then underwent a shortened version of the Trier Social Stress Task (TSST) [40,41] to induce psychological and physiological stress. The TSST consists of participants undergoing a public speaking task and arithmetic task in front of a panel of critically evaluative judges. The TSST has been shown to be a valid and reliable instrument for inducing physiological and psychological stress responses [42]. Immediately after the conclusion of the TSST, participants collected another saliva sample and self-reported their feelings of stress. For the next 5 minutes, participants were left alone in the room and interacted with their phone in a particular way depending on condition. Those in the *no phone* condition did not have their phone returned and were told to sit quietly for the next 5 minutes while the next portion of the study was prepared. Those in the *phone present* condition were given their phone but told "please do not use your phone for the remainder of the study." Those in *HRVB game* condition were told to open the Happify app, navigate to Breather, and "follow the instructions on the app." After the 5-minute phone manipulation period, the researcher returned to the room and instructed the participant to continue answering a series of questionnaires. Twenty minutes after completion of the TSST, a third saliva sample was collected. Forty minutes after the completion of the TSST, a fourth saliva sample was collected. At the conclusion of the study, the researcher and both judges debriefed the participant.

Measures

Demographics and Covariates

Demographic information and potential covariates, including age, sex, ethnicity, socioeconomic status, perceived psychological stress, measures of daily phone use, time since waking, use of hormonal contraceptives, and caffeine intake, were collected via self-report.

Self-Reported Stress

Participants were asked to indicate, "How stressed do you feel right now?" on a visual analog scale from 1-100. This simple one-item scale has been shown to be valid and reliable for assessing perceptions of acute stress [43]. Self-reported stress was assessed at three time points (baseline, post-TSST, +20 minutes recovery).

Physiological Stress

Salivary cortisol and sAA were both collected to provide a broad assessment of the physiological stress response. Since these salivary biomarkers indicate activity of different physiological stress systems and have different secretion times, the inclusion of both gives us a more comprehensive understanding of stress effects. Salivary cortisol and sAA were collected using a passive drool technique with polypropylene cryovial salivettes at four time points. The first three samples were assayed for sAA, and the last three samples were assayed for cortisol to accommodate for the differing secretion times of each analyte (ie, an approximately 20-minute lag time for salivary cortisol secretion into saliva compared to immediate secretion of sAA) and ensure that the collection timing aligned to capture measures of baseline, post-TSST, and +20 minute recovery time points. Experimental sessions were conducted in the afternoon (between 1 PM to 6 PM) to account for the diurnal rhythm of sAA and cortisol.

Salivettes were stored at –80°C until batch analysis at the end of data collection at the laboratory of the Institute for Interdisciplinary Salivary Bioscience Research (University of California Irvine, Irvine, California). Before assaying, the samples were thawed for an hour to return them to room temperature. For cortisol, all samples were assayed in duplicate by using an expanded-range high-sensitivity salivary cortisol enzyme immunoassay kit (Salimetrics, LLC, State College, Pennsylvania). The assay range of sensitivity was 0.007 to 3.0 μg/dL, and the average intraassay coefficient of variation was 5.5%. For sAA, samples were tested in duplicate using a commercially available kinetic enzyme reaction assay kit (Salimetrics, LLC). The assay range of sensitivity was 0.4-400 U/mL, and the average intraassay coefficient of variation was 3.3.%.

Analytic Strategy

All dependent variables (self-reported stress, sAA, and cortisol) were checked for skewness and kurtosis and transformed accordingly. No transformation was performed for values of self-reported stress. Values of sAA were moderately skewed, and a square root transformation was used to transform the values to approximate a normal distribution. Values of cortisol were moderately skewed, and a logarithmic transformation was used to transform the values to approximate a normal distribution. Outlying values above or below three SDs from the mean were removed. No outliers were removed for self-reported stress, three outliers (2.1%) were removed for sAA, and six outliers (2.8%) were removed for cortisol.

Models controlled for covariates that were significantly associated with the dependent variable. Sex, time since waking, and baseline cortisol were associated with cortisol recovery and were therefore controlled for in cortisol analyses. Baseline sAA was associated with sAA recovery and was therefore controlled for in sAA analyses. Baseline self-reported stress was associated with self-reported stress recovery and was therefore controlled for in self-reported stress analyses.

Independent sample t tests were used to conduct manipulation checks and ensure that exposure to the TSST reliably increased self-reported stress, sAA, and cortisol from baseline (time 1) to post-TSST stress (time 2). Repeated-measures mixed analysis of covariance was used to analyze the effect of condition on each dependent variable. Since the phone manipulation occurred after the TSST, analyses focused on differences in recovery and therefore used time (post-TSST stress at time 2 and +20 minute recovery at time 3) as the within-subject factor. Condition was included as a between-subject factor, and appropriate covariates were controlled for depending on the outcome of interest. Baseline values were controlled for to provide a more conservative and unbiased estimate of between-subject differences in composite recovery values [44]. Post-hoc comparisons were conducted to examine specific pairwise differences when a significant effect of condition was found.

Results

Manipulation Checks

Analysis of sAA from baseline (mean 85.31, SD 59.40) to post-TSST (mean 128.58, SD 88.53) revealed that participants displayed significant increases in sAA following the TSST (t_{276}=–4.78, $P<.001$). In addition, analysis of cortisol from baseline (mean 0.21, SD 0.11) to post-TSST (mean 0.34, SD 0.24) showed that participants displayed significant increases in cortisol following the TSST, (t_{271}=-5.99, $P<.001$). Finally, analysis of self-reported stress from baseline (mean 22.73, SD 21.48) to post-TSST (mean 47.31, SD 29.37) demonstrated that participants displayed significant increases in self-reported stress following the TSST (t_{276}=–7.96, $P<.001$). These results indicate that TSST reliably increased psychological and physiological stress.

Differences in Salivary Alpha Amylase Recovery Between Conditions

Between-subject comparisons indicated that there was a significant main effect of condition on sAA recovery ($F_{2,133}$=3.78, $P=.03$; *no phone*: mean 10.078, SE 0.271; *phone present*: mean 10.007, SE 0.280; *HRVB game*: mean 9.132, SE 0.266). Post-hoc comparisons revealed that those in the *HRVB game* condition displayed significantly less sAA during recovery than those in the *no phone* condition (t_{93}=2.48, $P=.02$) and the *phone present* condition (t_{90}=2.26, $P=.03$). The *no phone* and *phone present* conditions did not differ (t_{88}=–0.19, $P=.85$; Multimedia Appendix 2). Within-subject analyses revealed that there was no significant main effect of time for sAA recovery ($F_{1,133}$=.003, $P=.96$) and no significant interaction between time and condition ($F_{2,133}$=0.081, $P=.92$).

Differences in Cortisol Recovery Between Conditions

Although the cortisol levels declined during recovery for all conditions, there was no main effect of condition on cortisol recovery ($F_{2,126}$=1.19, P=.31).

Differences in Self-Reported Stress Recovery Between Conditions

Although self-reported stress declined during recovery for all conditions, there was no main effect of condition on self-reported stress recovery ($F_{2,133}$=1.42, P=.24).

Discussion

Principal Findings

In this study, we examined the effect of using or having a smartphone on psychological and physiological stress reduction during recovery. We found that those who used an HRVB training app exhibited the lowest levels of sAA during recovery. Specifically, those in the *HRVB game* condition released less sAA during recovery than those who had their phones present or had no phone at all. These results indicate that engaging in a brief 5-minute HRVB training session on a smartphone can effectively reduce stress-related sympathetic activity, as assessed by levels of sAA. Although the magnitude of the effect for the change in sAA was only small to medium (η^2=0.05) [45], it was similar to previous studies [16,46]. The sAA findings are particularly important because high levels of sAA are associated with a range of deleterious health-related outcomes such as asthma, frequency of illness, and chronic fatigue [47-49]; therefore, lower levels of sAA are desirable from a health perspective. Our findings provide health-related information about the use of mHealth interventions on a smartphone. Since delayed physiological recovery can be predictive of risk for long-term health issues [50], we can infer that using a serious game such as Breather when recovering from a stressful experience may provide long-term health benefits.

Our study design did not allow us to conclusively determine the mechanisms responsible for the stress-buffering effect, but there are several possibilities for why Breather effectively aided in stress recovery. The most obvious explanation is that undergoing HRVB training increases parasympathetic activity, which is typically inversely related to sympathetic indicators such as sAA. Thus, the low levels of sAA for those in the *HRVB game* condition may be indicative of direct physiological alterations induced by the use of Breather. In addition, psychological factors may have played a role in explaining sAA recovery. The simple distraction induced by diverting cognitive attention away from ruminating thoughts about the stressor may have positively contributed to the effects. Furthermore, parasympathetic activity has been associated with increases in positive valence and low arousal emotions such as calmn [24], which suggests that feelings of calm may have also played a role in stress recovery. Finally, it is possible that the ability to monitor stress responses via the visual interface of Breather increased perceptions of control, which subsequently alleviated feelings of stress. This is due to the fact that acute stress is often induced by a perceived lack of control [51], and when that perceived control is increased, it can inhibit autonomic arousal

[52]. It should be noted that no mechanisms can be determined for the lack of cortisol and self-reported stress, as we did not find significant effects on these measures. Future studies should further investigate the mechanisms for why HRVB delivered via a smartphone influences stress recovery.

Interestingly, those who had their phones present during stress recovery did not glean any additional stress-buffering benefits beyond those with no phone. Previous work has demonstrated that having a phone present, but not using it, leads to steep declines in sAA during physiological recovery from a stressor [16]. We failed to replicate this outcome. In the study by Hunter et al [16], participants had their phones with them while undergoing the stressor. In this study, participants only had their phones immediately after the study. This difference in timing implies that it may be helpful to have a phone present while experiencing a stressor, but it provides little to no benefit when present during recovery. In addition, a phone may serve as a "digital security blanket" in mildly stressful situations like social exclusion but may not exert similarly beneficial effects under more potent stressors such as the TSST.

Limitations

There are several limitations that limit the generalizability of these results. First, our sample is not representative of the population at large. The majority of our participants were healthy young Asian women, all of whom were iPhone users and college educated. Since we drew our sample from a university population, our participants were likely wealthier, younger, more dependent on their phone, and more educated than the average person. Thus, these conclusions cannot be extrapolated to all populations.

In addition, the effectiveness of Breather for influencing sAA recovery compared to the other conditions may have been statistically limited by differences in baseline values. Those in the *HRVB game* condition had significantly lower levels of sAA at baseline. These differences may have been due to a methodological inconsistency, as the individuals in the *HRVB game* condition had a slightly different experience during the baseline period before undergoing the TSST; they spent approximately 2 minutes receiving training on the HRVB app. Per methodological recommendations, these baseline values were controlled for to provide a more conservative and unbiased estimate of between-subject differences in recovery [43]. Without the inclusion of baseline sAA as a covariate in the models, there would have been greater statistical differences between *HRVB game* use and the other conditions. Although this statistical decision does limit the magnitude of our sAA recovery findings, these differences in baseline raise an interesting point about the ways in which Breather influences reactions to a stressor. If, indeed, the brief training period reduced baseline sAA and sAA reactivity to the stressor, then using an HRVB serious game like Breather could possibly be an effective method for buffering stress reactivity as well as recovery and may be an advisable activity to engage in prior to a major stressor. Before any recommendations can be made, future studies should explore the optimal timing for HRVB implementation and determine whether it is most effective before or after a stressor.

Additionally, the effectiveness of Breather may have been hindered by the way in which the participant interacted with the app and understood the directions. During this training period, information about the purpose of using this app (eg, this activity makes you more relaxed and less stressed) was hidden from the participant in order to reduce demand characteristics and maintain internal validity across conditions. However, the success of biofeedback training hinges on the individuals' perception that they are actively controlling their physiological functions in an effort to reduce stress [8]. Without this understanding about the purpose of the activity, the biofeedback exercise was likely less effective for the participants. In these ways, methodological constraints may have led to a more conservative effect of Breather compared to the other conditions.

Furthermore, user error issues that occurred within the app may have limited the effectiveness of Breather. The program requires the user's finger to be placed very precisely on the light sensor to monitor heart rate change. It is sometimes difficult to maintain this position, and warnings pop up on the screen each time a finger is placed incorrectly. Based on participant feedback, these warnings made individuals feel as if they were performing poorly, which may have induced further stress rather than alleviate it. To investigate whether user error played a role, adherence to finger placement was assessed using metrics provided from the app's database. Data showed that users had their fingers placed correctly for approximately 96% of the time; however, that still means that for 4% of the session, they were getting warnings telling them, "please place your finger on the sensor." This may have been bothersome and unduly reduced the effectiveness of Breather, which is important to consider in future efficacy tests and real-life applications.

When considering a more comprehensive assessment of stress, conclusions from this study must be tempered by the lack of significant group differences for salivary cortisol and self-reported stress. Based on these discrepancies, we can only conclude that the HRVB game had a targeted effect on autonomic nervous system recovery as opposed to a general effect on all types of biological and psychological stress recovery. These inconsistencies in stress outcomes may be due to a variety of reasons. First, cortisol and sAA represent activity in different arms of the stress system and are not correlated at a 1:1 level [36,53]. Numerous studies have discovered significant sAA results, but not cortisol, during stress recovery [37,38]. The one study that examined both biomarkers in the context of phone usage only found significant sAA effects [16]. Our findings indicate that the HRVB training had a more robust impact on autonomic nervous system activity (indicated by sAA measures), which makes sense because HRVB training specifically targets fluctuations in cardiovascular activity that is intricately tied to autonomic activity [36]. In addition, the intervention period was short (about 5 minutes), which may not have been enough time to impact cortisol, often viewed as a chronic stress marker with a delayed release [53]. The discrepancy between self-reported and physiological stress is quite common in studies that assess both constructs [17,34,42] and one of the reasons many researchers argue for the importance of assessing both when contemplating health relevance of stress or psychological outcomes [54]. Additionally, there is substantial variation based on individual factors, such as demographics, in the association of subjective and objective measures of stress [55]. Furthermore, studies examining the convergence of self-report and physiological measures of stress have found that the assessments are highly correlated during the TSST, but not before or after [42]. Thus, it is not surprising that sAA was the only metric that yielded significant results. The significant sAA finding provides valuable information about how a HRVB training game via an mHealth app may aid in stress recovery; however, future studies should consider a wider range of health-related outcomes.

Finally, it should be acknowledged that the scope of this study did not allow us to conclusively determine whether using the HRVB app was more or less effective than performing other actions on one's phone. Although past research is mixed on how phone use influences stress recovery [17,18], there is great potential for future researchers to explore how unstructured phone use (eg, listening to music, browsing social media) could impact physiological and psychological stress. Given the wide variety of potential ways in which people can use their phones, future studies should further investigate the effects of various types of phone interactions on stress recovery.

Conclusions and Implications

Based on these results, one can conclude that completing HRVB training on an app such as Happify may be a practical and effective strategy for reducing acute physiological stress. It is often not feasible to use a smartphone to buffer stress while undergoing a stressor, but it is practical and ecologically valid to use a phone immediately after one has experienced a stressful experience. Our smartphones are conveniently with us at most times, and thus, we have this effective stress-reducing tool at our disposal anytime and anywhere we need it. To further examine how smartphones can aid in stress recovery, future research should investigate the mechanisms underlying how a gamified stress-reducing app may buffer stress and how it compares to other ways of using a phone. This will inform future interventions and provide recommendations for the development of other stress-buffering tools that can be delivered through smartphone apps.

Results such as these are beginning to change the narrative about the effect of smartphones on our well-being. Although it is important to recognize the deleterious effects of these devices on our lives, it may be even more critical to recognize the positive potential of smartphones and begin to develop and use technology in ways that augment well-being. Instead of simply hoping that individuals use technology in a beneficial manner, it is imperative that the hardware and software are designed in a way that facilitates positive behavior, thoughts, and interactions. Designing tools that take advantage of the technological affordances and ubiquity of smartphones to put stress-reducing tools in the palm of one's hand is a promising strategy for finding ways for smartphones to maximize well-being.

Conflicts of Interest

ACP and ALW are full-time employees of Happify Health. However, all study design decisions, data collection, and analyses were performed by JFH, MSO, and SDP at the University of California, Irvine.

Multimedia Appendix 1

A brief demo video of Breather on the Happify app.

[MP4 File (MP4 Video), 157169 KB - games_v7i4e15974_app1.mp4]

Multimedia Appendix 2

Levels of salivary alpha amylase during recovery by experimental conditions.

[PNG File , 27 KB - games_v7i4e15974_app2.png]

References

1. Ralph BCW, Thomson DR, Cheyne JA, Smilek D. Media multitasking and failures of attention in everyday life. Psychol Res 2014 Sep;78(5):661-669. [doi: 10.1007/s00426-013-0523-7] [Medline: 24178629]

2. Roberts JA, Yaya LHP, Manolis C. The invisible addiction: cell-phone activities and addiction among male and female college students. J Behav Addict 2014 Dec;3(4):254-265 [FREE Full text] [doi: 10.1556/JBA.3.2014.015] [Medline: 25595966]

3. Rosen LD, Lim AF, Felt J, Carrier LM, Cheever NA, Lara-Ruiz JM, et al. Media and technology use predicts ill-being among children, preteens and teenagers independent of the negative health impacts of exercise and eating habits. Comput Human Behav 2014 Jun;35:364-375 [FREE Full text] [doi: 10.1016/j.chb.2014.01.036] [Medline: 25717216]

4. Twenge JM, Joiner TE, Rogers ML, Martin GN. Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time. Clinical Psychological Science 2017 Nov 14;6(1):3-17. [doi: 10.1177/2167702617723376]

5. McEwen BS, Gianaros PJ. Stress- and allostasis-induced brain plasticity. Annu Rev Med 2011;62:431-445 [FREE Full text] [doi: 10.1146/annurev-med-052209-100430] [Medline: 20707675]

6. McEwen BS. Stress, adaptation, and disease. Allostasis and allostatic load. Ann N Y Acad Sci 1998 May 01;840:33-44. [doi: 10.1111/j.1749-6632.1998.tb09546.x] [Medline: 9629234]

7. Goldenberg DL, Kaplan KH, Nadeau MG, Brodeur C, Smith S, Schmid CH. A Controlled Study of a Stress-Reduction, Cognitive-Behavioral Treatment Program in Fibromyalgia. Journal of Musculoskeletal Pain 2010 Jan 16;2(2):53-66. [doi: 10.1300/j094v02n02_05]

8. Moss D. Heart rate variability and biofeedback. Psychophysiology Today: The Magazine for Mind-Body Medicine 2004;1:4-11.

9. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]

10. Ward AF, Duke K, Gneezy A, Bos MW. Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity. Journal of the Association for Consumer Research 2017 Apr;2(2):140-154. [doi: 10.1086/691462]

11. Misra S, Cheng L, Genevie J, Yuan M. The iPhone Effect. Environment and Behavior 2014 Jul;48(2):275-298. [doi: 10.1177/0013916514539755]

12. Horrey WJ, Wickens CD. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Hum Factors 2006;48(1):196-205. [doi: 10.1518/001872006776412135] [Medline: 16696268]

13. Gerin W, Davidson KW, Christenfeld NJS, Goyal T, Schwartz JE. The role of angry rumination and distraction in blood pressure recovery from emotional arousal. Psychosom Med 2006;68(1):64-72. [doi: 10.1097/01.psy.0000195747.12404.aa] [Medline: 16449413]

14. Master SL, Eisenberger NI, Taylor SE, Naliboff BD, Shirinyan D, Lieberman MD. A picture's worth: partner photographs reduce experimentally induced pain. Psychol Sci 2009 Nov;20(11):1316-1318. [doi: 10.1111/j.1467-9280.2009.02444.x] [Medline: 19788531]

15. Brown SL, Nesse RM, Vinokur AD, Smith DM. Providing social support may be more beneficial than receiving it: results from a prospective study of mortality. Psychol Sci 2003 Jul;14(4):320-327. [Medline: 12807404]

16. Hunter JF, Hooker ED, Rohleder N, Pressman SD. The Use of Smartphones as a Digital Security Blanket: The Influence of Phone Use and Availability on Psychological and Physiological Responses to Social Exclusion. Psychosom Med 2018 May;80(4):345-352. [doi: 10.1097/PSY.0000000000000568] [Medline: 29521885]

XSL•FO

RenderX

17. Rus HM, Tiemensma J. Social media as a shield: Facebook buffers acute stress. Physiol Behav 2018 Mar 01;185:46-54. [doi: 10.1016/j.physbeh.2017.12.021] [Medline: 29273455]

18. Rus HM, Tiemensma J. Social Media under the Skin: Facebook Use after Acute Stress Impairs Cortisol Recovery. Front Psychol 2017 Sep 19;8. [doi: 10.3389/fpsyg.2017.01609]

19. Hooker ED, Campos B, Pressman SD. It just takes a text: Partner text messages can reduce cardiovascular responses to stress in females. Computers in Human Behavior 2018 Jul;84:485-492. [doi: 10.1016/j.chb.2018.02.033]

20. Lin I, Peper E. Psychophysiological patterns during cell phone text messaging: a preliminary study. Appl Psychophysiol Biofeedback 2009 Mar;34(1):53-57. [doi: 10.1007/s10484-009-9078-1] [Medline: 19199025]

21. Afifi TD, Zamanzadeh N, Harrison K, Acevedo Callejas M. WIRED: The impact of media and technology use on stress (cortisol) and inflammation (interleukin IL-6) in fast paced families. Computers in Human Behavior 2018 Apr;81:265-273. [doi: 10.1016/j.chb.2017.12.010]

22. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]

23. Parks AC, Williams AL, Tugade MM, Hokes KE, Honomichl RD, Zilca RD. Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: A randomized controlled trial. Intnl J Wellbeing 2018 Dec 08;8(2):22-67. [doi: 10.5502/ijw.v8i2.745]

24. Lehrer PM, Gevirtz R. Heart rate variability biofeedback: how and why does it work? Front Psychol 2014;5:756 [FREE Full text] [doi: 10.3389/fpsyg.2014.00756] [Medline: 25101026]

25. Lacey BC, Lacey JI. Two-way communication between the heart and the brain. Significance of time within the cardiac cycle. Am Psychol 1978 Feb;33(2):99-113. [Medline: 637402]

26. Karemaker J, Lie KI. Heart rate variability: a telltale of health or disease. Eur Heart J 2000 Mar 15;21(6):435-437. [doi: 10.1053/euhj.1999.1969] [Medline: 10681483]

27. Del Pozo JM, Gevirtz RN, Scher B, Guarneri E. Biofeedback treatment increases heart rate variability in patients with known coronary artery disease. Am Heart J 2004 Mar;147(3):E11. [doi: 10.1016/j.ahj.2003.08.013] [Medline: 14999213]

28. Kleiger RE, Miller J, Bigger J, Moss AJ. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. The American Journal of Cardiology 1987 Feb;59(4):256-262. [doi: 10.1016/0002-9149(87)90795-8]

29. Dillon A, Kelly M, Robertson IH, Robertson DA. Smartphone Applications Utilizing Biofeedback Can Aid Stress Reduction. Front Psychol 2016;7:832 [FREE Full text] [doi: 10.3389/fpsyg.2016.00832] [Medline: 27378963]

30. Prinsloo GE, Rauch HGL, Lambert MI, Muench F, Noakes TD, Derman WE. The effect of short duration heart rate variability (HRV) biofeedback on cognitive performance during laboratory induced cognitive stress. Appl Cognit Psychol 2010 Oct 17;25(5):792-801. [doi: 10.1002/acp.1750]

31. Stein P. Washington University in St. Louis, School of Medicine. 2018. Validation of the Happify breather biofeedback exercise to track heart rate variability using an optical sensor. URL: https://happify.com/health/results/HLS-05.pdf [accessed 2019-11-11]

32. Vaschillo E, Lehrer P, Rishe N, Konstantinov M. Heart rate variability biofeedback as a method for assessing baroreflex function: A preliminary study of resonance in the cardiovascular system. Applied Psychophysiology and Biofeedback 2002;27(1). [doi: 10.1023/A:1014587304314]

33. Robins R, Fraley R, Krueger R. The self-report method. In: Handbook Of Research Methods In Personality Psychology. New York: The Guilford Press; 2019.

34. Campbell J, Ehlert U. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? Psychoneuroendocrinology 2012 Aug;37(8):1111-1134. [doi: 10.1016/j.psyneuen.2011.12.010] [Medline: 22260938]

35. Hellhammer DH, Wüst S, Kudielka BM. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology 2009 Feb;34(2):163-171. [doi: 10.1016/j.psyneuen.2008.10.026]

36. Nater UM, Rohleder N. Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: current state of research. Psychoneuroendocrinology 2009 May;34(4):486-496. [doi: 10.1016/j.psyneuen.2009.01.014] [Medline: 19249160]

37. Payne LA, Hibel LC, Granger DA, Tsao JCI, Zeltzer LK. Relationship of Salivary Alpha Amylase and Cortisol to Social Anxiety in Healthy Children Undergoing Laboratory Pain Tasks. J Child Adolesc Behav 2014;2 [FREE Full text] [doi: 10.4172/jcalb.1000129] [Medline: 25525630]

38. van Veen JF, van Vliet IM, Derijk RH, van Pelt J, Mertens B, Zitman FG. Elevated alpha-amylase but not cortisol in generalized social anxiety disorder. Psychoneuroendocrinology 2008 Nov;33(10):1313-1321. [doi: 10.1016/j.psyneuen.2008.07.004] [Medline: 18757137]

39. Winslow BD, Chadderdon GL, Dechmerowski SJ, Jones DL, Kalkstein S, Greene JL, et al. Development and Clinical Evaluation of an mHealth Application for Stress Management. Front Psychiatry 2016 Jul;7:130 [FREE Full text] [doi: 10.3389/fpsyt.2016.00130] [Medline: 27507949]

40. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993;28(1-2):76-81. [doi: 10.1159/000119004] [Medline: 8255414]

XSL•FO
RenderX

41. Allen AP, Kennedy PJ, Dockray S, Cryan JF, Dinan TG, Clarke G. The Trier Social Stress Test: Principles and practice. Neurobiol Stress 2017 Feb;6:113-126 [FREE Full text] [doi: 10.1016/j.ynstr.2016.11.001] [Medline: 28229114]

42. Hellhammer J, Schubert M. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 2012 Jan;37(1):119-124. [doi: 10.1016/j.psyneuen.2011.05.012] [Medline: 21689890]

43. Lesage F, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. Occup Med (Lond) 2012 Dec;62(8):600-605 [FREE Full text] [doi: 10.1093/occmed/kqs140] [Medline: 22965867]

44. Overall JE, Doyle SR. Implications of chance baseline differences in repeated measurement designs. J Biopharm Stat 1994 Jul;4(2):199-216. [doi: 10.1080/10543409408835083] [Medline: 7951275]

45. Richardson JT. Eta squared and partial eta squared as measures of effect size in educational research. Educational Research Review 2011 Jan;6(2):135-147. [doi: 10.1016/j.edurev.2010.12.001]

46. Stroud LR, Foster E, Papandonatos GD, Handwerger K, Granger DA, Kivlighan KT, et al. Stress response and the adolescent transition: performance versus peer rejection stressors. Dev Psychopathol 2009;21(1):47-68 [FREE Full text] [doi: 10.1017/S0954579409000042] [Medline: 19144222]

47. Granger DA, Kivlighan KT, Blair C, El-Sheikh M, Mize J, Lisonbee JA, et al. Integrating the measurement of salivary α-amylase into studies of child health, development, and social relationships. Journal of Social and Personal Relationships 2016 Jun 29;23(2):267-290. [doi: 10.1177/0265407506062479]

48. Granger DA, Kivlighan KT, el-Sheikh M, Gordis EB, Stroud LR. Salivary alpha-amylase in biobehavioral research: recent developments and applications. Ann N Y Acad Sci 2007 Mar;1098:122-144. [doi: 10.1196/annals.1384.008] [Medline: 17332070]

49. Wolf JM, Nicholls E, Chen E. Chronic stress, salivary cortisol, and alpha-amylase in children with asthma and healthy children. Biol Psychol 2008 Apr;78(1):20-28. [doi: 10.1016/j.biopsycho.2007.12.004] [Medline: 18243483]

50. Stewart JC, France CR. Cardiovascular recovery from stress predicts longitudinal changes in blood pressure. Biol Psychol 2001 Nov;58(2):105-120. [doi: 10.1016/s0301-0511(01)00105-3] [Medline: 11600240]

51. Eysenck M. Anxiety: The cognitive perspective. London, UK: Psychology Press; 2013.

52. Leotti LA, Iyengar SS, Ochsner KN. Born to choose: the origins and value of the need for control. Trends Cogn Sci 2010 Oct;14(10):457-463 [FREE Full text] [doi: 10.1016/j.tics.2010.08.001] [Medline: 20817592]

53. Engert V, Vogel S, Efanov SI, Duchesne A, Corbo V, Ali N, et al. Investigation into the cross-correlation of salivary cortisol and alpha-amylase responses to psychological stress. Psychoneuroendocrinology 2011 Oct;36(9):1294-1302. [doi: 10.1016/j.psyneuen.2011.02.018] [Medline: 21470780]

54. Diener E, Pressman SD, Hunter J, Delgadillo-Chase D. If, Why, and When Subjective Well-Being Influences Health, and Future Needed Research. Appl Psychol Health Well-Being 2017 Jul 14;9(2):133-167. [doi: 10.1111/aphw.12090]

55. Föhr T, Tolvanen A, Myllymäki T, Järvelä-Reijonen E, Rantala S, Korpela R, et al. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. J Occup Med Toxicol 2015;10:39 [FREE Full text] [doi: 10.1186/s12995-015-0081-6] [Medline: 26504485]

Abbreviations

HRV: heart rate variability
HRVB: heart rate variability biofeedback
mHealth: mobile health
mHealth: mobile health
sAA: salivary alpha amylase
TSST: Trier Social Stress Task

XSL•FO
RenderX

<u>Original Paper</u>

Development of a Mobile Game to Influence Behavior Determinants of HIV Service Uptake Among Key Populations in the Philippines: User-Centered Design Process

Charlotte Hemingway[1], BA; Emmanuel S Baja[2], SCD, BS, MSc; Godafreda V Dalmacion[2], BSc, MSc, MD; Paul Mark B Medina[2], BS, MSc, PhD; Ernest Genesis Guevara[2], BS; Tyrone Reden Sy[2], MPH, MA; Russell Dacombe[1], MSc; Claire Dormann[3], MSc, PhD; Miriam Taegtmeyer[1,4], BM, MA, ChB, FRCP, PhD

[1]Department of International Public Health, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

[2]Institute of Clinical Epidemiology, National Institutes of Health, University of the Philippines, Manila, Philippines

[3]Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

[4]Tropical Infectious Diseases Unit, Royal Liverpool University Hospital, Liverpool, United Kingdom

Corresponding Author:
Charlotte Hemingway, BA
Department of International Public Health
Liverpool School of Tropical Medicine
Pembroke Place
Liverpool, L35QA
United Kingdom
Phone: 44 7942118311
Email: charlotte.hemingway@lstmed.ac.uk

Abstract

Background: Opportunities in digital distribution place mobile games as a promising platform for games for health. However, designing a game that can compete in the saturated mobile games market and deliver persuasive health messages can feel like an insurmountable challenge. Although user-centered design is widely advocated, factors such as the user's subject domain expertise, budget constraints, and poor data collection methods can restrict the benefits of user involvement.

Objective: This study aimed to develop a playable and acceptable game for health, targeted at young key populations in the Philippines.

Methods: Authors identified a range of user-centered design methods to be used in tandem from published literature. The resulting design process involved a phased approach, with 40 primary and secondary users engaged during the initial ideation and prototype testing stages. Selected methods included participatory design workshops, playtests, playability heuristics, and focus group discussions. Subject domain experts were allocated roles in the development team. Data were analyzed using a framework approach. Conceptual frameworks in health intervention acceptability and game design guided the analysis. In-game events were captured through the Unity Analytics service to monitor uptake and game use over a 12-month period.

Results: Early user involvement revealed a strong desire for online multiplayer gameplay, yet most reported that access to this type of game was restricted because of technical and economic constraints. A role-playing game (RPG) with combat elements was identified as a very appealing gameplay style. Findings guided us to a game that could be played offline and that blended RPG elements, such as narrative and turn-based combat, with match-3 puzzles. Although the game received a positive response during playtests, gameplay was at times perceived as repetitive and predicted to only appeal to casual gamers. Knowledge transfer was predominantly achieved through interpretation of the game's narrative, highlighting this as an important design element. Uptake of the game was positive; between December 1, 2017, and December 1, 2018, 3325 unique device installs were reported globally. Game metrics provided evidence of adoption by young key populations in the Philippines. Game uptake and use were substantially higher in regions where direct engagement with target users took place.

Conclusions: User-centered design activities supported the identification of important contextual requirements. Multiple data collection methods enabled triangulation of findings to mediate the inherent biases of the different techniques. Game acceptance is dependent on the ability of the development team to implement design solutions that address the needs and desires of target users. If target users are expected to develop design solutions, they must have adequate expertise and a significant role within

XSL•FO
RenderX

the development team. Facilitating meaningful partnerships between health professionals, the games industry, and end users will support the games for health industry as it matures.

Introduction

The Philippines' HIV Epidemic

The Philippines has the fastest growing HIV epidemic in the Southeast Asia region. The dominant mode of transmission reported in the Philippines is sexual contact among males who have sex with males (MSM) and transgender women (TGW). As of 2017, the Joint United Nations Programme on HIV and AIDS reported the HIV prevalence rate at 0.1% for general adult population (aged 15-49 years), 0.3% among young men (aged 10-24 years), 4.9% among MSM, and 1.7% among TGW [1]. Specific data for TGW are limited [2]; however, a cross-sectional study in Cebu City in 2015 showed an 11.8% HIV prevalence rate in this group [3]. Cases are forecast to triple in the next 10 years, with the majority of new infections among young MSM (aged 15-24 years) [4]. Since the first reported case in 1984, the National HIV/AIDS and Antiretroviral Therapy Registry of the Philippines has confirmed 59,135 cases as of September 2018, of which 28% were aged between 15 and 24 years and 51% were aged between 25 and 34 years at the time of diagnosis. The proportion of HIV-positive cases in the 15- to 24-years age group has almost tripled in the last 10 years [5].

Young key populations worldwide pose a complex public health challenge; transmission rates are high, diagnosis is often delayed, and linkage to care and treatment is poor among those found to be infected [6]. Barriers to HIV services vary across contexts. In the Philippines, intrapersonal and social barriers exist alongside health system and economic barriers. Recurring themes associated with barriers to HIV services in the literature include low perceived risk of HIV infection; fear of losing access or status in important social spheres as a repercussion of accessing HIV services; lack of awareness or negative perceptions of treatment; belief that clinics do not provide confidential or private services; and restricted access to testing services because of time constraints, economic constraints, and legal constraints for those aged under 18 years [7-11].

Why Play Mobile Games?

As the HIV epidemic worsens, social changes are occurring for young people in the Philippines. Increased online connectivity, a growing economy, and prevalent mobile device use has changed the way people spend their time and socialize. This change has generated new possibilities for the public health sector to deliver targeted health messages [12]. Utilizing technology already popular among adolescents and young adults may provide access to individuals who do not otherwise engage with traditional forms of HIV education and advocacy [13]. Mobile games, in particular, offer a promising platform to address the knowledge gaps, perceptions, social pressure, and self-stigma that deter young key populations from accessing

health services [14]. A recent meta-analysis of 54 digital games for healthy lifestyle promotion found small but significant effects on behavior, determinants, and clinical outcomes, demonstrating the potential benefits [15].

Mobile games contain structural elements that effectively engage users [16], and it is through these elements that mobile games could influence behavior determinants. Players can model health-related behavior and witness positive and negative outcomes within a safe environment. Well-designed narratives, integrated with the gameplay, can foster identification with the characters, thereby increasing a player's sense of personal risk or self-efficacy in overcoming barriers to HIV services [17]. Positive portrayals of characters living with HIV may help form beliefs that players can also remain or become their desired self after a positive diagnosis. Influence may also be found from the complex interplay between digital gaming and social behavior [18]. Even a single-player mobile game can trigger meaningful social interactions, from recommending a new game to offering advice on how to complete a challenge. For example, a player may recommend the game to a peer or family member who they believe could benefit from the health-related content. Narratives and characters within the game could also trigger meaningful conversations within social groups. Such use of the game could generate social pressure about health-related behavior.

A Need for User-Centered Design in Games for Health

Rittel and Webber [19] defined the term *wicked problem* as a design problem that cannot be solved in a stepwise problem-solving manner. Game development is riddled with wicked problems. They are inherently interactive, and any interactive component must be tested with users. Even a simple game has an interconnected system as its backbone; any change to 1 component of the game will have ramifications to all connected components. Furthermore, game development is bound by countless budget, technical, user, and market constraints. As a solution to wicked problems, some game developers turned to user-centered design [20]. User-centered design is an iterative process; using a range of research techniques, feedback is obtained from users at different stages of product development to ensure their needs and preferences are considered in the design [21]. This design approach was established in the software industry to identify and rectify usability issues and expanded for use in game design to evaluate experiential aspects, supporting the development of games that are both functional and fun. User involvement in game design may have benefits with regard to the effectiveness of the game, although the evidence is mixed when reviewed in the context of games for healthy lifestyle promotion [22]. The core principle behind this theory is that, for a game to be effective, it must be acceptable and appropriate. This can only be achieved through

consultation with target users, especially if the developers do not share the same characteristics as them. User involvement is also predicted to improve user adoption and enable developers to critically reflect on the value and consequences of a game when vulnerable groups are involved [23]. Thus, user-centered design methods must encompass interpretation and emotional response to health-related content in the game as well as evaluating usability.

As the purpose of our game expanded beyond entertainment, we developed a theory of behavior change (Figure 1) to guide the design process [24]. Context-specific learning objectives were generated through a review of the literature and formative research with target users and HIV service providers in the regions of Davao and Manila. Data were collected through focus group discussions (FGDs) with HIV services providers, in-depth interviews with MSM and TGW, and an online survey on enablers and barriers to HIV services. Findings from the formative research are not included in detail in this paper. The framework references modifying factors relevant to the Philippines' HIV testing context such as health system and legal and economic constraints. Given that the target audience for the game will be young MSM and TGW, we do not expect the game to impact modifying factors but note that these are important considerations in the evaluation of the game's perceived effectiveness.

Figure 1. Integrative model of predictive behaviour and predicted game influence.
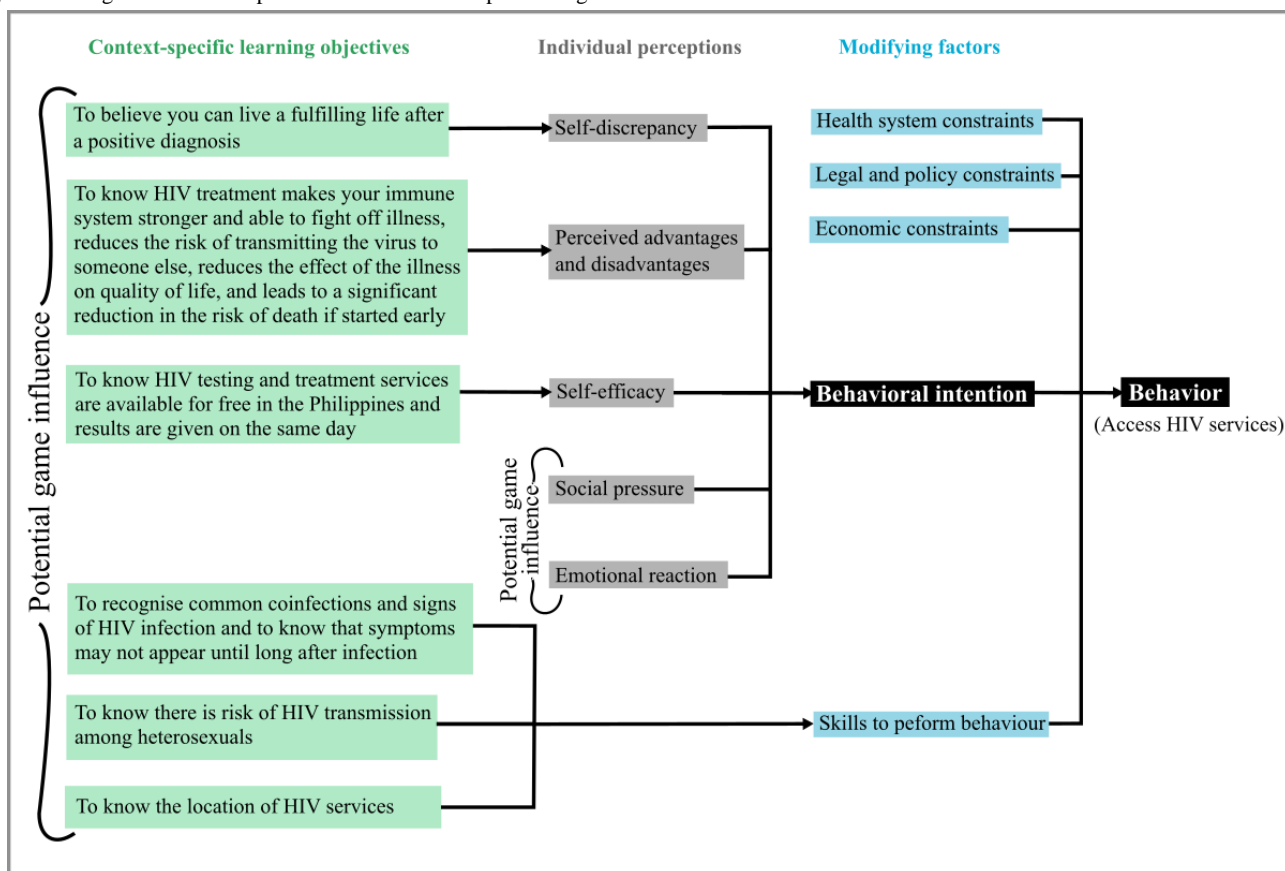
In this paper, we describe the user-centered design process for a mobile game titled *Battle in the Blood* and explore the effect of user-centered design techniques on the game's acceptability as a health intervention and uptake among target users.

Methods

Overview

Battle in the Blood was developed from 2016 to 2018 through a collaborative effort with experts from the United Kingdom and the Philippines. Final design choices were made by the game's coproducers: a British behavioral scientist specializing in game design for health system benefit (CH); the director of an independent game development company, with over 10 years industry experience, based in Scotland; and a Filipino clinical epidemiologist (EB). The process was divided into 3 phases and aligned with development milestones for the game (Figure 2). Internal playtests were frequently conducted by members of the project team and their immediate networks. Change requests and bugs were recorded and shared with the development team using Google Docs.

Figure 2. Development process.
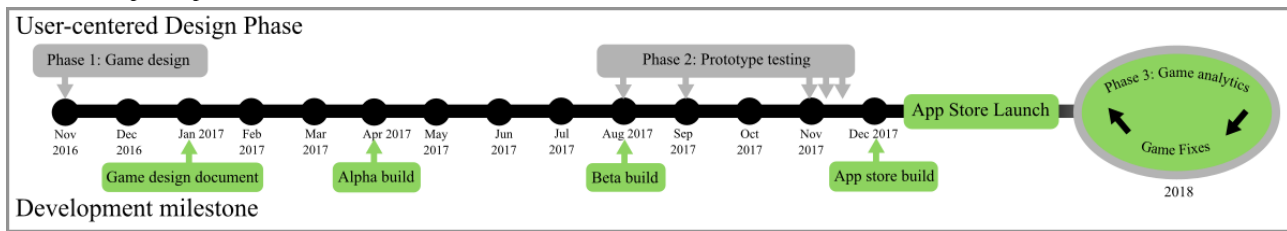
Selection of User Centered Design Methods

A literature review was conducted to identify a range of user-centered design methods to be used in a combined approach. Methods were selected based on their fit to the question the development team wanted to answer and the available resources. The resulting design process involved a phased approach, with primary and secondary users engaged during the initial ideation and prototype testing stages. Resources were prioritized to capture important contextual requirements and evaluate educational and experiential aspects of the game. Selected methods included participatory design workshops [25], extended playtests [20], playability heuristics [26], FGDs, and game analytics.

As the methods for each phase were distinct, recruitment, data collection, and analysis have been described for each. Data were collected in the cities of Manila and Davao. The study was part of a much larger project, which included assessment of a rapid diagnostic algorithm for HIV, that was being piloted in these cities.

Phase 1: Game Design

To determine important contextual requirements and user preferences, design workshops with Filipino gamers were conducted during the initial specification and ideation stage. Subject experts in clinical practices for HIV and the Philippines HIV epidemic were allocated informant roles in the game development team.

Recruitment

Filipino individuals older than 18 years who regularly played digital games or were involved in game development or electronic sports, regardless of sexual identity, were invited to participatory design workshops via social networking sites, including Facebook and Steam forums. It was theorized that the game would need to be appealing and accessible to a range of gamer types and that gamer type would not be dictated by sexual identity. A total of 18 participants were divided into 2 groups (11 and 7). Participants were aged between 21 and 30 years and were a mixed group of MSM and non-MSM.

Data Collection

Group sessions were facilitated by UK and Filipino researchers (CH, EB, EG, and JD), in which (1) participants responded to questions by creating a human scatter graph indicating their level of agreement to different statements by their physical proximity to the statement placed on the floor, creating an instant visual of the group's perception and experience of mobile games and games for health; (2) group discussions were held on design and technical enablers and barriers to digital gaming; and (3)

participants were divided into teams (maximum 6 participants in a team) to develop pitches for games to promote HIV services, which they then presented and discussed. HIV clinic counselors were present at both workshops to answer any questions the participants had regarding HIV and provided information on HIV service provision in the Philippines. The session was audio recorded and transcribed, and photos were taken to record the human scatter graphs and visuals from the game pitches.

Analysis

Transcripts and session outputs were analyzed using a framework approach [27] to identify a set of recommended game features. A 4-day workshop was conducted with the full development team to translate the list of recommended game features into a game design document. Game features were also connected to the context-specific learning objectives outlined in Figure 1.

Phase 2: Prototype Testing

Playtesting sessions and FGDs were conducted to assess the acceptability and playability of the beta game build.

Recruitment

Participants were older than 18 years and self-identified as MSM. This study aimed to ensure that the narratives and art style were appealing, relatable, and inoffensive to the target users. Participants were recruited using social networking sites, including Facebook, Grindr, Growlr, and PlanetRomeo, known as popular networking sites for the MSM and TGW community. Peer counselors from HIV testing services were also recruited as they were predicted to be an important user group for the game.

Overall, 5 FGDs and playtesting sessions involving a total of 22 participants were conducted between August 20, 2017, and November 18, 2017; 4 sessions were conducted in Manila (17 participants, all MSM) and 1 in Davao (5 participants, all peer counselors and MSM). Game changes and bug fixes were implemented between each testing session.

Data Collection

Sessions were facilitated by UK and Filipino researchers (CH, EB, EG, and JD). User testing was divided into 3 activities: (1) participants' screens and faces were video recorded as they played the game for 30 to 45 min, (2) all animations in the game were played on a big screen and a short discussion was held by participants for each one, and (3) a group discussion was conducted on the perceived acceptability of the game.

Analysis

Analysis of data from the game testing sessions was divided into 2 parts. The first part utilized playability heuristics to

identify and fix design flaws. Playability heuristics are a set of qualities by which a game's engagement and usability can be assessed. They are typically used by game developers and professional game testers. We adapted an existing list of playability heuristics for mobile games [26] to use as a coding framework to analyze the video recordings and transcripts (Figure 3). Screen recordings were primarily used to identify technical and usability issues, whereas transcripts were more conducive to recognizing different experiential aspects. For example, expressions of confusion by the player when they lost a level, especially when followed by repeated failed attempts,

could be marked as a violation of gameplay heuristic GP1, "the game provides clear goals" (Figure 3). This is an example of what is termed a playability violation. Playability reports were developed independently by 2 researchers (CH and EG). The reports described each playability violation and gave recommendations for game improvements. The reports were compared and discussed by the game development team before agreeing on a final list of game changes. Where feasible, changes were implemented before the next game testing session. Software bugs and technical issues were also recorded, and fixes were implemented.

Figure 3. Playability heuristics for mobile game for health. UI: user interface.
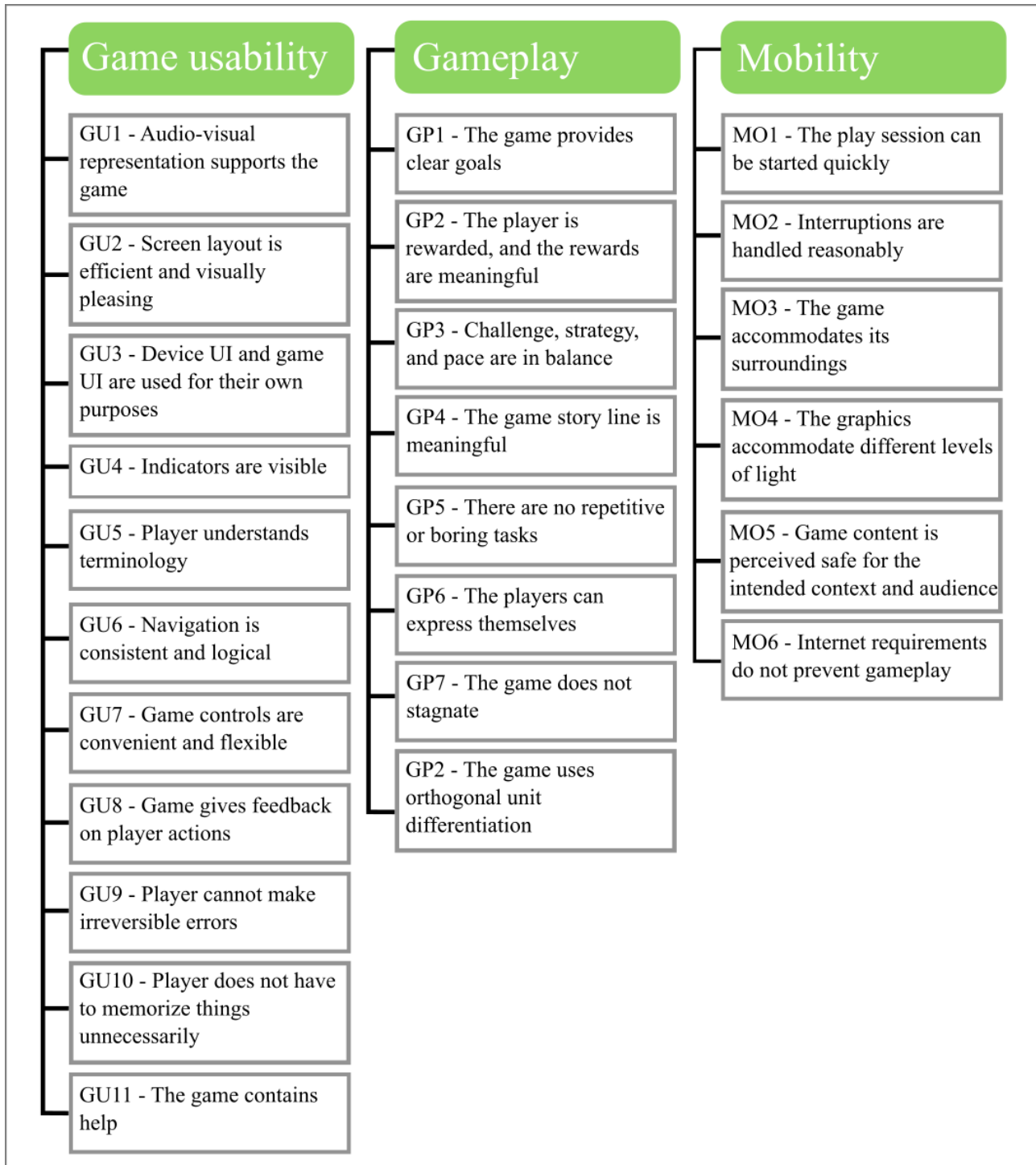
The second part of the data analysis utilized the transcription of the group discussions to assess the acceptability of the game as a health intervention. Data analysis was informed by a general inductive approach, aligning emerging themes identified in the transcripts with the following predetermined constructs of health intervention acceptability: affective attitude, burden, ethicality, intervention coherence, opportunity costs, perceived effectiveness, and self-efficacy [28].

Quality Assurance: Phases 1 and 2

All qualitative data collected were translated into English for analysis (EG), and the translation was checked for accuracy by members of the research team. Analysis was led by a UK researcher (CH) with regular consultation and input from the full project team to improve quality and depth. All design choices resulting from the data were reviewed by the development team before implementation.

Phase 3: Game Analytics

Recruitment

Game use data were obtained from users who installed the game and gave permission for the app to access device storage. Geolocation data were obtained from users who granted access to their devices' location data and played the game with GPS switched on. Information about why data were being gathered and reassurances that all data would be kept anonymous were presented on the app store page, and in the game when access was requested.

The game was made available on the Apple App Store and Google Play on November 27, 2017, and was officially launched during World AIDS Day celebrations in the Davao Region on December 1, 2017. Marketing events for the game included local television appearances by project staff, printed and online news articles, exhibitions at health- and game-related conferences in the National Capital Region and Davao Region, social media advertisements on Facebook and Twitter, and posters displayed in 4 HIV testing and counseling clinics (3 in Manila and 1 in Davao). The game analytics span a 12-month period from December 1, 2017, to December 1, 2018.

Data Collection

Data were automatically gathered from devices with the game installed through the Unity Analytics service, version 2017.1 by Unity Technologies. For data to be sent from the device, the user must have opened the game at least once while the device was connected to the internet. Events are stored locally on the device when the game is played offline and sent the next time the game is opened and the device is connected. All data points were stored in the Unity Data Store. The game metrics (Table 1) were exported and converted into a readable format for use in statistical analysis software.

Table 1. Game metrics.

Segment	Metrics
Active player metrics	Daily active users; monthly active users; and new users (unique device installs)
Session metrics	Sessions per day; sessions per user; total daily playing time; and total playing time per active user
Retention metrics	Day 1, 7, and 30 retention
Platform segment	Total Android and iOS users
Custom segments	Level events—reports each time a level is completed or failed, and the players end score; link event—reports the click-through rate of all in-game links to external information; question events—reports the response to all in-game questions, including a set of demographic questions; geolocation per session—reports the device's current location at the start of a session to 2 decimal places; and app version

Analysis

Descriptive statistical analysis was conducted in Microsoft Excel, version 1812. Spatial analysis was conducted in QGIS, version 3.2.3.

Ethical Assurance: All Phases

Legal advice was sought to ensure that management of game analytics complied with the 2018 European Union General Data Protection Regulation.

Ethical approval for the project was obtained from the ethics committees at the University of the Philippines College of Medicine and Liverpool School of Tropical Medicine (research protocol 16-017).

Results

Phase 1: Game Design

Direction on game elements to support health-related behavior change first emerged through the human scatter graph activity. Figure 4 demonstrates that the 2 groups had varied perceptions on the effectiveness of games for health-related behavior change. Participants were questioned as to why they had chosen their place on the graph. Perceived limitations of the game's effectiveness were centered around beliefs that changes in knowledge would not be sufficient to change behavior and that the use of technical language would make the game's content inaccessible, as illustrated by the following quote:

Figure 4. Human scatter graph on perceptions of the effectiveness of games for health behavior change.
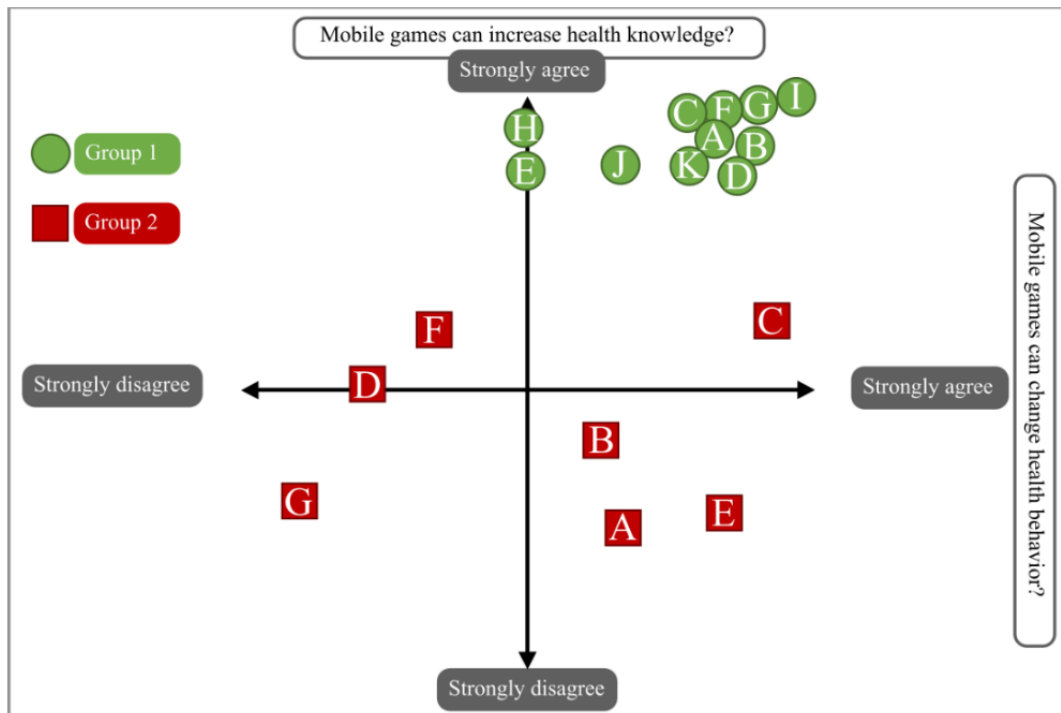
I'm still not convinced with the knowledge part because, as with some health conditions, even with some health benefits and stuff, there will be some things that a player will not understand, especially if it's scientific jargon...but I do believe that behavior can change especially if it's immersive and it is an experience that will change your perspective on very different matters. [Group 2: Respondent B]

Facilitators to behavior change centered around the use of immersive experiences that could change perspectives and generate social interactions between users, as illustrated by the following quote:

Because whenever I'm playing games, they would often ask me "Hi [name]! What's that game? Is that available in iOS? Is it available in the Android Play Store?" Right after I told them about it, they would go check out the game and download it. So, it's not just me but also them that learns from the game. [Group 1: Respondent J]

During group discussion, the game's narrative was identified as an important feature. If done well, it fostered an identification with the game characters and motivated the players to overcome challenges in the gameplay to witness the story unfold. The narrative was identified as the logical place to communicate why it was important to know one's HIV status. Participants felt that the game should not provide technical information on HIV or HIV services but should focus on telling an emotionally driven story and provide the player choice over the narrative direction.

When the groups were divided into teams (2 teams in each group) to pitch their HIV advocacy games, 3 out of the 4 teams presented a similar concept of a hero sent on a quest to fight or evade the HIV virus, where story and role-playing game (RPG)

elements intertwined and where reality and fantasy existed in the same space.

Multiplayer gameplay was reported to be a strong motivator for repeated gameplay and was perceived to play a role in the effectiveness of the game, as illustrated by the following quote:

...they say that no man is an island. So, I think it would affect the behaviour of the person if there is a community that pushes you. Someone playing alone will say "I'm just alone, nobody will care if I do this or do that." But if you know somebody else is pushing you, it will affect your behaviour, it drives you as a person. [Group 1: Respondent I]

However, participants also reported that access to these types of games was restricted because of the requirement of a stable internet connection:

I have external and internal factors why I leave the game. The external factor is whenever the game requires internet connection because the internet, the wireless data in the Philippines is not that good. [Group 1: Respondent E]

This presented a dilemma for the game design. On the one hand, multiplayer gameplay was a very desirable game style among the participants as they could participate in discussions about HIV, behavior, and personal values and coconstruct their own narratives about a desired future. This was further evidenced by the market success of multiplayer mobile games in the Philippines. On the other hand, we did not have the resources to develop and maintain an online multiplayer platform, and known issues around internet connectivity would restrict access to the game.

Combining Feasibility With Desire

Findings from the phase 1 game design workshops directed the game's overarching narrative and the integration of gameplay and storytelling. Players take on the role of the protagonist, entering the blood stream in an antiretroviral (ARV) pill capsule and battling anthropomorphic viruses, bacteria, and cancer cells using a weaponized mechanical suit. The gameplay combines match-3 puzzles with turn-based combat; players connect icons on a puzzle board that represent condoms, ARVs, healthy living, health care, and time. Connecting the icons builds up the player's defense and attack status during the rounds of combat. The match-3 game style was selected because of the availability of prebuilt game assets and source code, which substantially reduced the development time and enabled the team to allocate time to the custom animations.

The gameplay is segmented with a series of 8 animated stories about people living with HIV that the player helps by progressing through the game. The difficulty of the levels and the types of enemy units are connected with the story line. For example, if the character in the story line is diagnosed with gonorrhea, an enemy unit representing gonococcus appears during the gameplay. In the game's final mission, the player is introduced to a character with AIDS in critical condition; it is revealed that he has never been tested for HIV, treatment fails, and the character dies. Throughout the game, the player is awarded with fragments that combine to form an anting-anting, a traditional Filipino amulet believed to have magical powers. This amulet allows the player to travel back in time and change how the story ends for this character by encouraging him to undergo a HIV test.

The game can be played and completed offline. Online features include a global leader board, ranking players by their total score, and hyperlinks to websites containing information on HIV and HIV services in the Philippines. The file size of the game was also restricted to 53.4 MB on Android and 70 MB on iOS to reduce the risk of failed download attempts from the app stores. This in turn had ramifications for animations and sounds in the game, both of which can substantially increase the file size. To maintain a small file size, the number of sound effects and music tracks in the game was limited and animated stories were presented in a dynamic 2-dimensional comic book style, where static images moved across the screen to give the scenes depth and movement.

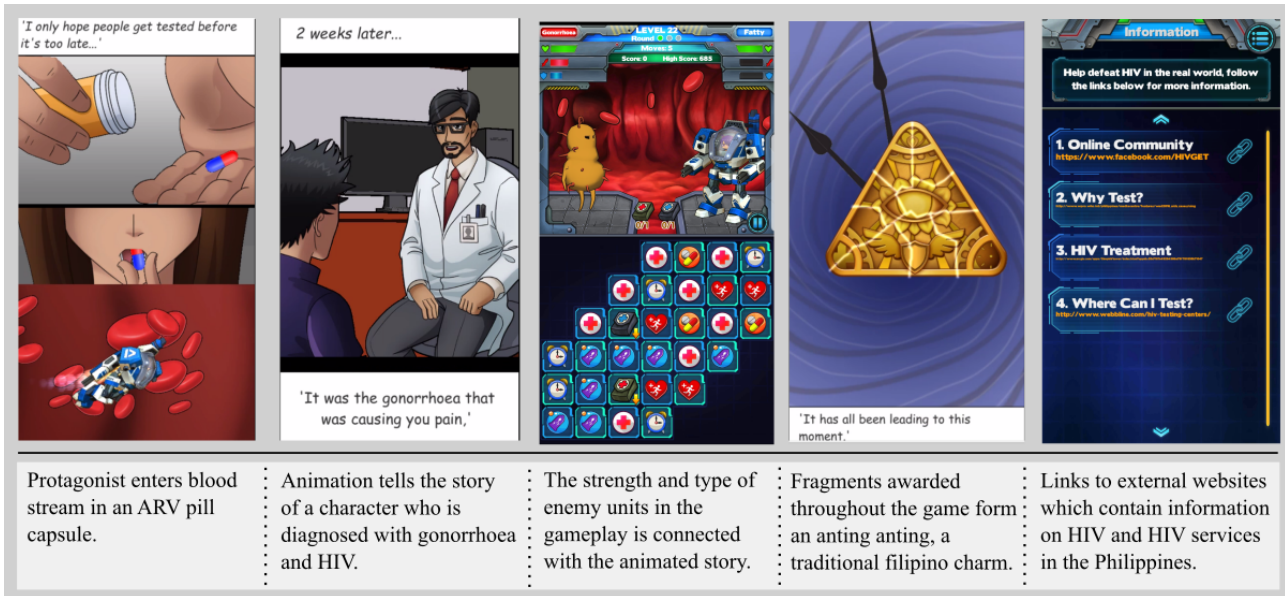### Integration of Context-Specific Learning Objectives

Table 2 details the design choices made in relation to the context-specific learning objectives outlined in Figure 1.

Figure 5 uses screenshots from the game to illustrate some of the design choices made. A table summarizing the phase 1 design process (Multimedia Appendix 1) and the game design document, version 0.15, (Multimedia Appendix 2) are included as Multimedia Appendices.

Table 2. Theory of behavior change design choices.

Context-specific learning objective	Game design element
To believe that you can live a fulfilling life after a positive diagnosis	The animated stories depict people living with HIV pursuing their ambitions, being socially active, or spending time with their family. For example, a transgender character enters a beauty pageant and, as she takes her medication in the dressing room, the text reads, "I fought to be my true gender, I can win this fight too." This particular story line was selected because of the cultural relevance pageantry has in the Philippines and the great esteem in which Filipinos hold the contestants.
To know that HIV treatment makes your immune system stronger and more able to fight off illness; reduces the risk of transmitting the virus to someone else; reduces the effect of the illness on quality of life; and, if started early, leads to a significant reduction in the risk of death	Antiretroviral tablets feature in every animated story except the one where the character dies. Treatment is also shown to improve the physical appearance of the characters in the later missions. Players can earn an extra life during gameplay by answering a multiple-choice question. Some questions address knowledge on treatment effects, and feedback containing the right answer is given. Hyperlinks to websites containing information on HIV treatment are included throughout the game, and the click-through rate is measured.
To know that HIV testing and treatment services are available for free in the Philippines and that results are given on the same day	In the animated stories, the words "Free HIV tests" appear in the background of all the clinic waiting room scenes. The rapid testing procedure is depicted in the first 3 animations. Multiple-choice questions address the availability of free testing and the time taken for the client to receive their results in the Philippines. Hyperlinks to websites containing information on HIV services in the Philippines are included throughout the game, and the click-through rate is measured.
To recognize common coinfections and signs of HIV infection and to know that symptoms may not appear until long after infection	Enemy units in the game include representations of gonorrhea, tuberculosis, hepatitis B, herpes, and cancer. Common signs of HIV infection are depicted in the animated stories, and some characters show no symptoms.
To know that there is a risk of HIV transmission among heterosexuals	The first animated story depicts a heterosexual couple. In several other animated stories, the character's sexuality is ambiguous.
To know the location of HIV services	A hyperlink to a website containing the contact information and location of all testing sites in the Philippines is featured throughout the game.

Figure 5. Screenshots Battle in the Blood v1.4.

Protagonist enters blood stream in an ARV pill capsule.	Animation tells the story of a character who is diagnosed with gonorrhoea and HIV.	The strength and type of enemy units in the gameplay is connected with the animated story.	Fragments awarded throughout the game form an anting anting, a traditional filipino charm.	Links to external websites which contain information on HIV and HIV services in the Philippines.

Phase 2: Prototype Testing

Table 3 summarizes the final agreed list of playability violations and the game changes made as a result.

Learning Achieved Predominantly Through Interpretation of the Animated Narratives

Participants repeatedly drew inferences between the presence of ARVs and the positive outcome in the short-animated stories, recognizing the unique benefits of treatment for each character, as illustrated by the following quote:

> *Ever since the guy was given the ARVs, his fear of him having STDs or socializing with people is now gone, and he was able to not be afraid to connect with other people.* [GameTester_MSM]

Table 3. Playability violations and game changes.

Playability heuristic	Evidence of violation	Game change
GP5: There are no repetitive or boring tasks	Participants felt bored and disengaged with the animation when the first half of the story line was repeated. Video footage demonstrated that players did not realize they could skip the recap by tapping anywhere on the screen.	On the second playthrough, a title screen appears with the text "Previously on *Battle in the Blood*…Skip?"
GU1: Audio-visual representations support the game	Players reported the icons as unintuitive because the color of the icons did not correspond with the combat status bars that they effected. Video footage showed players taking decisions based solely on the length of the chain, not the color of the icons, and becoming frustrated when they repeatedly lost a level.	The design of the icons was adjusted to correspond visually with the attack and defense bars.
GU11: The game contains help	Participants did not feel that the onboarding was comprehensive enough, and new mechanics in the game were not explained.	When a new game mechanic is triggered, a dialog box appears that explains the new mechanic using a small amount of text and images. This information can also be accessed via the game's menu and via the level-pause screen.
GP6: The players can express themselves	Participants stated that they did not like the mechanical suit as they had spent time customizing the avatar but rarely saw it in the game as most of it was covered.	Avatar appears animated in the level-complete or -fail screen.
GP2: The player is rewarded, and the rewards are meaningful; GP5: There are no repetitive or boring tasks	Participants stated that they wanted to be able to earn in-game currency by completing the levels, which could be spent on upgrading or customizing their character. They also felt this would make the gameplay feel less repetitive.	Implementing an in-game currency system and custom character upgrades would have required additional resources and delayed the planned launch of the game. The change request was logged but not implemented for the pilot.

In addition to the inferred benefits of HIV treatment, participants also identified messages about the consequences of delayed access to HIV services. One participant reported that the game had changed his personal understanding of when to undergo a HIV test, as shown by the following conversation:

For me it's better to test at the earliest, so you can know if you're positive and prevent it earlier. [PhaseTwo_GameTester_MSM]

So, before you played the game, you just know that you need to get tested when? [Interviewer]

Only if you're already having symptoms. [PhaseTwo_GameTester_MSM]

Participants interpreted the stories based on their current knowledge and personal values. In 1 case, this led to the message of the story being repeatedly challenged:

...he was told that he has Gonorrhoea and he was then given pills to cure it and the HIV treatment. With that, he told himself that this could save his life and then could continue living his life normally. And then he was seen to be at the bar again. But it was somehow an off for me because it gave me the impression that just because there's treatment, you can go on with your promiscuous activity all your life. [PhaseTwo_GameTester_MSM]

Some counterarguments were expressed because of preexisting knowledge, namely, a lack of awareness that ARVs can reduce the viral load to the point where the individual is no longer infectious and social norms in which sexual promiscuity is perceived as an undesirable trait. Most participants concluded that the character in this particular animation was sexually promiscuous because the story centered around a nightclub, and he was diagnosed with HIV and gonorrhea. In this case, the narrative appeared to challenge stigmatizing perceptions around sexual promiscuity and its association with HIV.

Although participants felt that the health-related messages were reinforced in the turn-based combat levels, most of the learning outcomes were achieved through interpretation of the animated narratives and the discussions they triggered. A potential concern is that the animations were perceived to be disconnected from, and less engaging than, the gameplay. During the prototype testing session, all in-game animations were played in sequence through a projector, and participants intently watched and discussed each animation. Interpretation of the narrative may be significantly different by a user casually playing the game alone, with no direct incentive to focus on and analyze the content.

Identification With Game Characters

Animated narratives where the players shared a set of commonalities with the character diagnosed with HIV elicited a stronger emotional reaction than those where the characters did not share similar demographics or behaviors. The story of a young MSM who resorts to solicitation to afford to play games at an internet café was reported to be the most impactful story line by almost all game testers. This was likely because of the recruitment strategy, as most game testers were part of the young MSM community and had a strong interest in gaming:

Well it seems relatable for me because I'm a gamer. And I have a friend who does that thing sometimes. And thankfully he hasn't got any HIV. It's relatable to me because well, I know those who did, I

experienced one but thankfully I haven't got HIV. [GameTester_MSM]

The game follows the stories of 8 different characters diagnosed with HIV. Each character is unique in its sexuality, gender, appearance, and behavior. Therefore, a user is likely to empathize with or relate to some but not all characters in the game. As game characters are presented in a predetermined sequence, some users may never witness a character that they share commonalities with, reducing the potential impact of the in-game narrative for certain users based on their progress in the game. The representation of different sexualities and genders was well received by participants as they acknowledged that it could help to address misconceptions that HIV only effects MSM, indicating that the design flaw was the delivery of the narratives rather than the inclusion of multiple character types.

Gameplay Perceived as Enjoyable but Potentially Limited in Appeal

Despite a positive response from the participants, in which gameplay was described as "enjoyable, addictive, and challenging," there were indications that the game would have a limited appeal. In the first instance, participants had very low expectations for the game and expressed surprise at playing a game comparable with commercial games on the app stores. In 1 FGD, participants identified 2 categories of gamers, casual and hardcore, and discussed which category the game would appeal to and why. A casual gamer was perceived to be someone who occasionally engaged with digital games as a form of distraction, whereas a hardcore gamer's life would revolve around digital games. Most agreed that the game would have little to no appeal for hardcore gamers. For casual gamers, the consensus was that the game would be appealing but required an effective marketing and deployment strategy to overcome competition from similar games. Offline gameplay was perceived to be a strong motivator, regardless of gamer type, as it would enable players to alleviate boredom when they were unable to access features on their devices that required an internet connection. Marketing *Battle in the Blood* as a game about HIV was perceived to have 2 effects on the appeal; in most cases, it would arouse curiosity or tap into a desire to learn, whereas in some cases, it could be off-putting. Participants felt that the game's marketing should also be targeted toward parents and should offer information on the age appropriateness of the content.

Phase 3: Game Analytics

Game analytics reported 3325 unique device installs globally during a 12-month period. The game received an average of 10 installs per day. Installs peaked during active and incentivized marketing events; the maximum number of installs in 1 day was 367 at the time of the World AIDS Day launch event in Davao.

Unity Analytics provided a more accurate report of game uptake and reach than the app store services because of the common practice of using third-party apps, such as SHAREit, to access the game via Bluetooth and avoid data costs. Installs via this method would not have been captured by the analytics services provided by the Apple App Store or Google Play at the time of the study.

Regional Game Uptake

Geolocation data were collected for at least one play session from 50.77% (1688/3325) of users. Of the users from whom geolocation data were collected, 85.36% (1441/1688) were located in the Philippines (Figure 6). Game use was concentrated in urbanized areas. Uptake of the game was highest in the National Capital Region and Davao Region. Furthermore, 54.01% (1796/3325) of the users reported their age, gender, and gender of sexual partners; of these users, 28.06% (504/1796) reported as being sexually active MSM. Within the sexually active MSM category, 47.8% (241/504) reported their age as between 25 and 34 years, 26.1% (132/504) as between 20 and 24 years, and 7.1% (36/504) as between 10 and 19 years.

Figure 6. Cluster point map of the location of Battle in the Blood installations in the Philippines. Points clustered at 4mm. Number in circle represents total number of users which installed Battle in the Blood in that location. Total number of user records from the game analytics that contain geolocation data and have played the game in the Philippines=1441. Region level shape file from Humanitarian Data Exchange.

Overall, 2 factors are believed to have contributed to the substantially higher uptake among target users in the regions of Davao and Manila. First, several promotional exhibitions for the game were held by project staff in these regions. The exhibitions often included prize giveaways for those who downloaded the game, indicating a dependency on these types of active and incentivized marketing events. Self-reported data indicated that marketing events targeting adolescent users were potentially lacking. Second, user-centered design activities were conducted in these regions with influencers in both the gaming and HIV community, and it is likely that they also played a role in promoting the game.

Low Rates of Game Completion

As of December 1, 2018, 14.98% (498/3325) of users were reported to have completed level 45 and 4.00% (133/3325) were reported to have completed level 90, the last level in the game. Factors believed to have contributed to the reported low rates of game completion are the gameplay being perceived as repetitive and having limited appeal and missing data because of the provision of offline gameplay.

Additional factors impacting uptake and progression will be explored further through interviews with end users and follow-up interviews with participants involved in the design process.

Identification of Access Issue on Android Devices

Game analytics and app store reviews were monitored throughout the 12-month period. Monitoring activities supported the identification of an access issue on Android devices, which was related to app permissions managed by the Google Play store. A fix was launched on November 5, 2018. Between October 5, 2018, and November 4, 2018, the average number of daily active users was 16, and this increased to 40 in the month following the fix.

Discussion

Establishing Design Solutions Through User-Centered Design

User-centered design methods supported the identification of contextual requirements for the game [22]. Qualitative methods provided a deep understanding of the important factors that both motivated and enabled gameplay, which are likely to have contributed to the reported uptake and use of the game. The human scatter graph method and game pitches enabled participants to reflect on their views and the views of others, which, in turn, ignited valuable discussion on the qualities the game required to both be appealing and deliver persuasive health messages. Although findings from phase 1 influenced the creative direction of the game, each design choice made by the development team was prefixed with a discussion on what was feasible within the predetermined budget and development time frame. For example, during phase 1, an action RPG in which the player could control their character's strength and traits was found to be highly desirable, but adequate resources were not available to implement a stable and balanced underlying mathematical model for this feature. In this case, the design workshops supported the identification of a wicked problem,

the solution for which was informed by the experience and best judgment of the game's coproducers [19]. The reported low rates of game completion indicated that optimal design solutions were not achieved by the game's producers.

User-Centered Design and Community Adoption

Results from the game analytics indicated that users' involvement in the game development process had a positive impact on uptake among target users. However, the extent of that impact cannot be determined from the available data. For example, although unique device installs were substantially higher in regions where user-centered design activities took place, the proportion of installs credited to such activities cannot be determined, especially with the presence of targeted and incentivized promotional events in those regions. Further studies are recommended to explore the correlation between user involvement and community adoption of games for health. The way in which target users involved in the design are credited may also be an important factor to consider.

Narratives in Games for Health

Storytelling has long been established as an effective means of attitude and behavior change [17], and this was reflected in the findings during the user-centered design process. This also highlights the importance of having an experienced narrative designer as part of the development team to work in partnership with the subject domain and behavioral experts. Assessment of the game's acceptability found that mobile game design that accommodates typical user behavior is not always conducive to effective storytelling. To be suitable for use in a public space, the animated stories had to be highly captivating to retain focus while also allowing the player to disengage without missing vital information. Although *Battle in the Blood* made some headway in delivering a narrative that was accessible and accommodated typical user behavior, a lack of interaction, an obscure link to the gameplay, delivery through linear episodes, and limited character identification may have rendered the animated stories ineffective in generating new knowledge and perceptions when the game is played outside a facilitated gameplay session. These design flaws could be addressed through the inclusion of additional branching narratives driven by player choice, forming a stronger link between the gameplay and cause and effect sequences in the narrative, inclusion of stories based on true events, allowing players to select narratives they are interested in, and facilitating group gameplay sessions where players are encouraged to discuss their interpretation of the narratives as part of the game's distribution strategy.

Multiple Data Collection Methods to Mediate Biases

Telemetry and geolocation data are widely used in the mobile games industry to improve app revenue. There are countless articles ranking key performance indicators by their value to app developers. In response to demand, a range of services are now available that process large volumes of data into actionable information with very low setup costs. In the case of the Unity Analytics service (version 2017.1), standard metrics can be obtained by toggling a switch in the Unity game engine. Although generating income may not be a primary goal, it could be argued that analytical services that capture data during

gameplay are currently underutilized in games for health development and evaluation [29]. In contrast to FGDs, telemetry and geolocation data collected through the game are not biased by the presence of a researcher; however, emotive responses to gameplay cannot be captured this way. Thus, it is important not to focus solely on 1 data collection technique but rather to triangulate results from a range of different data sources to mitigate bias while providing a more complete picture of game performance.

This study explored the effect of user involvement on health intervention acceptability and community adoption and described in detail the methods used. Further evaluation of the game has been conducted to explore the game's effect on knowledge, attitudes, and HIV service use among target users, the results of which will be published in a separate paper.

Limitations

As with all qualitative research, findings cannot be applied to the wider population with certainty. Perceptions from aged under 18 years and subjects outside Manila and Davao were not captured. Although the project team failed to recruit TGW during phases 1 and 2 of the game development process, they were included during the formative research stage, which informed the conceptual framework for the game (Figure 1). Feedback on how the transgender character in the game was portrayed was provided by TGW known to the research team in the Philippines in an informal capacity.

The authors note that the recruitment strategy and use of FGDs is likely to have resulted in a bias toward participants less encumbered by stigma toward their sexuality or HIV status. Given that recruitment advertisements stated that participants would be inputting into the design of an HIV advocacy game,

this may have also created a bias toward individuals with a vested interest in health advocacy and game design.

The timing for phase 2 protype testing was not ideal, which is believed to have impacted the changes implemented in the game design.

Certain terms used by the participants did not have a direct English translation, which, at times, led to ambiguity in the data during the translation process.

Conclusions

By involving users from the outset, the development team was guided toward narratives that were shown to be relatable and understandable and to gameplay that provided enjoyment. The resulting product is a game that is accessible, simple, and entertaining because of universally recognized game mechanics, small file size, and offline gameplay. At a minimum, games for health should involve target users in early stages of the design process as a relatively cost-effective method of cataloging important contextual requirements and user preferences. If target users are to be tasked with developing design solutions, then interaction with the development team must go beyond FGDs. For the target user to be a valued member of the game development team, they must have adequate expertise in design, a shared goal, and be properly credited and compensated for their contributions [22]. Increased reporting of design approaches and stronger collaboration among health professionals, the entertainment games industry, and end users will support the games for health industry as it matures. Restructuring projects to involve users before determining the development budget will enable design choices to be driven more by user requirements and less by what is feasible.

Authors' Contributions

CH analyzed the data and wrote the manuscript. CH and EB produced the game. All authors inputted into the study design, game design, data analysis, and manuscript. EG and TS coordinated the recruitment and data collection activities and transcribed and translated the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of phase 1 design process.
[DOCX File , 18 KB - games_v7i4e13695_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Battle in the Blood game design document v0.15.
[DOCX File , 3318 KB - games_v7i4e13695_app2.docx]

References

1. The Joint United Nations Programme on HIV/AIDS. UNAIDS Philippines Country Factsheet. URL: https://www.unaids.org/en/regionscountries/countries/philippines [accessed 2017-12-01]
2. Bumanglag KC. Enhancing HIV/AIDS surveillance in the Philippines to ensure the transgender population's visibility. Am J Public Health 2018 Jan;108(1):58-60. [doi: 10.2105/AJPH.2017.304158] [Medline: 29211538]
3. Department of Health website. 2015. Integrated HIV behavioral and serologic surveillance: 2015 fact sheet. URL: https://www.doh.gov.ph/sites/default/files/publications/Philippines_2015_IHBSS_Fact_Sheets_Nov2017_22JUN2016.pdf [accessed 2017-12-01]
4. Department of Health website. 2016. The State of the Philippine HIV Epidemic 2016 Facing Challenges, Forging Solutions. URL: https://www.doh.gov.ph/sites/default/files/publications/publication__nonSerials_State%20of%20HIV%20Epidemic%20in%20the%20Philippines.pdf [accessed 2017-12-01]
5. AIDS Data Hub. 2018. HIV / AIDS and ART Registry of the Philippines. URL: http://www.aidsdatahub.org/hivaids-and-art-registry-philippines-march-2018-epidemiology-bureau-department-health-2018 [accessed 2017-12-01]
6. Wong VJ, Murray KR, Phelps BR, Vermund SH, McCarraher DR. Adolescents, young people, and the 90-90-90 goals: a call to improve HIV testing and linkage to treatment. AIDS 2017 Jul 1;31(Suppl 3):S191-S194 [FREE Full text] [doi: 10.1097/QAD.0000000000001539] [Medline: 28665876]
7. Adia A, Bermudez A, Callahan M, Hernandez L, Imperial R, Operario D. 'An Evil Lurking Behind You': drivers, experiences, and consequences of HIV-related stigma among men who have sex with men with HIV in Manila, Philippines. AIDS Educ Prev 2018 Aug;30(4):322-334. [doi: 10.1521/aeap.2018.30.4.322] [Medline: 30148671]
8. van Wijngaarden JW, Ching AD, Settle E, van Griensven F, Cruz RC, Newman PA. 'I am not promiscuous enough!': exploring the low uptake of HIV testing by gay men and other men who have sex with men in Metro Manila, Philippines. PLoS One 2018;13(7):e0200256 [FREE Full text] [doi: 10.1371/journal.pone.0200256] [Medline: 29979766]
9. Canoy NA, Ofreneo MA. Struggling to care: a discursive-material analysis of negotiating agency among HIV-positive MSM. Health (London) 2017 Nov;21(6):575-594. [doi: 10.1177/1363459315628040] [Medline: 26851265]
10. Burki T. HIV in the Philippines. Lancet Infect Dis 2017 Jun;17(6):589-590. [doi: 10.1016/S1473-3099(17)30256-6] [Medline: 28555587]
11. Koirala S, Deuba K, Nampaisan O, Marrone G, Ekström AM, CAT-S group. Facilitators and barriers for retention in HIV care between testing and treatment in Asia-A study in Bangladesh, Indonesia, Lao, Nepal, Pakistan, Philippines and Vietnam. PLoS One 2017;12(5):e0176914 [FREE Full text] [doi: 10.1371/journal.pone.0176914] [Medline: 28459881]
12. KPMG Global. 2018. IT Report: Philippines Investment Guide. URL: https://home.kpmg/ph/en/home/insights/2017/12/it-report-philippines.html [accessed 2017-12-01]
13. Newzoo. 2017. The Filipino Gamer. URL: https://newzoo.com/insights/infographics/the-filipino-gamer/ [accessed 2017-08-20]
14. Hightow-Weidman LB, Muessig KE, Bauermeister J, Zhang C, LeGrand S. Youth, technology, and HIV: recent advances and future directions. Curr HIV/AIDS Rep 2015 Dec;12(4):500-515 [FREE Full text] [doi: 10.1007/s11904-015-0280-x] [Medline: 26385582]
15. DeSmet A, van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
16. Prensky M. Fun, play and games: what makes games engaging. In: Digital Game-Based Learning. United States: Paragon House; 2001:1-31.
17. Shen F, Han J. Effectiveness of entertainment education in communicating health information: a systematic review. Asian J Commun 2014;24(6):605-616. [doi: 10.1080/01292986.2014.927895]
18. Kort YA, IJsselsteijn WA, Poels K. Digital games as social presence technology : development of the Social Presence in Gaming Questionnaire (SPGQ). In: Presence 2007 : the 10th Annual International Workshop on Presence. Barcelona: Starlab; 2007:1-9.
19. Rittel HW, Webber MM. Dilemmas in a general theory of planning. Policy Sci 1973;4(2):155-169. [doi: 10.1007/BF01405730]
20. Pagulayan R, Keeker K, Fuller T, Wixon D, Romero R, Gunn DV. User-centered design in games. In: Sears A, Jacko JA, editors. Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Boca Raton: CRC Press; 2007:883-906.
21. Giacomin J. What is human centred design? Des J 2014;17(4):606-623. [doi: 10.2752/175630614X14056185480186]
22. DeSmet A, Thompson D, Baranowski T, Palmeira A, Verloigne M, de Bourdeaudhuij I. Is participatory design associated with the effectiveness of serious digital games for healthy lifestyle promotion? A meta-analysis. J Med Internet Res 2016 Apr 29;18(4):e94 [FREE Full text] [doi: 10.2196/jmir.4444] [Medline: 27129447]

XSL•FO
RenderX

23. Mayer I, Bekebrede G, Harteveld C, Warmelink H, Zhou Q, van Ruijven T, et al. The research and evaluation of serious games: toward a comprehensive methodology. Br J Educ Technol 2014;45(3):502-527. [doi: 10.1111/bjet.12067]

24. Conner M. Cognitive determinants of health behavior. In: Steptoe A, editor. Handbook Of Behavioral Medicine: Methods And Applications. New York, NY: Springer; 2010:19-30.

25. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]

26. Korhonen H, Koivisto EM. Playability Heuristics for Mobile Multi-player Games. In: Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts. 2007 Presented at: DIMEA'07; September 19-21, 2007; Perth, Australia p. 28-35.

27. Smith J, Firth J. Qualitative data analysis: the framework approach. Nurse Res 2011;18(2):52-62. [doi: 10.7748/nr2011.01.18.2.52.c8284] [Medline: 21319484]

28. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]

29. Smith SP, Blackmore K, Nesbitt K. A meta-analysis of data collection in serious games research. In: Loh CS, Sheng Y, Ifenthaler D, editors. Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement. Switzerland: Springer International Publishing; 2015:31-55.

Abbreviations

ARV: antiretroviral
FGD: focus group discussion
MSM: males who have sex with males
RPG: role-playing game
TGW: transgender women

XSL•FO
RenderX

Original Paper

Using Narrative Game Design to Increase Children's Physical Activity: Exploratory Thematic Analysis

Amy Shirong Lu[1], PhD; Melanie C Green[2], PhD; Debbe Thompson[3], PhD

[1]Health Technology Lab, College of Arts, Media, and Design, Bouvé College of Health Sciences, Northeastern University, Boston, MA, United States

[2]Department of Communication, State University of New York at Buffalo, Buffalo, NY, United States

[3]Children's Nutrition Research Center, Baylor College of Medicine, US Department of Agriculture, Houston, TX, United States

Corresponding Author:
Amy Shirong Lu, PhD
Health Technology Lab
College of Arts, Media, and Design, Bouvé College of Health Sciences
Northeastern University
360 Huntington Ave
212A Lake Hall
Boston, MA
United States
Phone: 1 617 373 7839
Email: a.lu@northeastern.edu

Abstract

Background: Physical activity is crucial for child obesity prevention and intervention. Narratives embedded in active games can increase children's physical activity.

Objective: Little is known about the narrative characteristics that would motivate children to exercise. We attempted to fill the gaps in understanding regarding narrative design for active video games.

Methods: In this exploratory study, four animated narratives of different genres were professionally generated. Children (N=41) between the ages of 8 and 12 years were interviewed to identify their preferences. Sessions were digitally recorded, transcribed, and analyzed using exploratory thematic analysis.

Results: Findings revealed that the children rated the dystopian science fiction story as their favorite across all weight, race, and gender groups. The physical activity-friendly narrative strategies included virtuous characters, extraordinary character actions, interesting plots, super powers, and engaging cliffhangers. Alternatively, information not related to physical activity, difficult-to-follow plot lines, passive protagonists, and repetitive narrative tropes were less appealing for physical activity.

Conclusions: This research provides preliminary evidence that narratives have characteristics that may increase child physical activity when playing active games. Future empirical studies should verify and test these design principles.

Introduction

Background

Childhood obesity is a growing epidemic in the United States and beyond [1]. In 2016, 340 million youth aged 5 to 19 years around the world were overweight or obese [2]. The obesity trend is increasing among many groups, with no ages seeing a decrease [3]. Obesity is a major risk factor for noncommunicable diseases such as cardiovascular disease, diabetes, and several types of cancers [4].

Physical activity (PA) is a critical strategy for child obesity prevention [5]. While it is recommended that children participate in 60 minutes or more of PA every day [6], less than one-quarter meet the recommendation [7]. Most PA interventions have not achieved long-term effects, with lack of access and motivation identified as key challenges [8].

Active video games (AVGs) require body movement to play [9] and may increase PA in an enjoyable manner [10]. With the international popularity of Dance Dance Revolution (Konami), Wii (Nintendo), and Pokémon Go (Niantic), AVG became

recognized as an independent genre. In the United States, approximately two-thirds of people play video games daily, and 70% of families have a child who plays video games [11]. All major consoles offer AVG devices [12]. AVG may also encourage PA among children in unsafe neighborhoods lacking safe outdoor alternatives [13].

Many AVGs, though, are perceived to be less enjoyable than sedentary games and are less likely to be played over time [14]. Most gamers do not complete their games [15] as they often lack long-term motivational appeal [16,17]. Novel and interdisciplinary approaches are needed to sustain AVG play [18].

Narratives, or stories, possess unique motivational properties that may encourage increased AVG play [19]. While narratives appear in some health games, few AVGs capable of achieving moderate to vigorous physical activity (MVPA) have incorporated them [20-22]. Recent studies have tested the ability of added narratives to induce PA through AVG play. An exploratory study that interspersed narrative cutscenes (ie, brief, animated narrative video clips) in an existing AVG without a narrative found that children aged 8 to 12 years in the narrative condition had 40% more objectively measured steps than their counterparts who played the nonnarrative version of the same game [23]. Similar results were found in another study involving college students; the narrative increased MVPA by 58% [24].

Both of the aforementioned studies featured narratives that included fantastical elements. For example, the first study used a mystery-themed story of a modern child absorbed into a game world trying to find the way back home [23]. The second featured an ordinary person fending off swarms of monsters to protect the world [24]. However, the question remains as to what narrative elements motivate children to engage in PA. Two exploratory studies provided preliminary insight: actions related to PA should be a central narrative element, a credible PA motivation should be incorporated, and characters should be likeable [25,26]. These studies, however, did not employ a theoretical foundation for analysis and largely consisted of anecdotal observations.

The elusive narrative characteristics that would best motivate PA among children should be explored with the guidance of interdisciplinary theories and in-depth analysis of children's feedback about the AVG narratives [27]. Here we attempt to fill this gap by exploring narrative design principles.

Narratives for Health Promotion in an Interactive Technology Age

A narrative is defined as any two or more events arranged in a temporal order [28]. Characters and plot are primary components of narratives. The characters are a major structural property and driving force [29], serving as an internal source of beliefs [30]. As described by social cognitive theory [31,32], characters also function as role models. The plot plays a pivotal role by organizing events into a logically unfolding sequence or temporal order [33].

Narratives have strong potential for health promotion by influencing cognition, affect, and behavior through transportation [34,35]. Transportation is narrative's unique

immersive quality that enables the suspension of disbelief, instills vivid personal experience, and creates deep affection for characters [36].

The advancement of AVG systems has created numerous opportunities for narratives to be better woven with the digital technologies and, in addition, helps amplify the foundational mechanisms of narrative persuasion. For example, the interactive nature of AVGs enhances player engagement with the characters through increased character identification, interpersonal attraction, and parasocial interaction. An appealing plot helps players to go beyond the spectator level to become an active participant and fosters greater engagement with the plot through gameplay. Therefore, an engaging AVG with appealing characters and plots could induce a strong intrinsic motivation to play exergames by reducing cognitive load [37,38] through immersive qualities [39], engendering a positive and powerful arousal and attention [40], enhancing character identification [41], and absorbing players in a story world [42]. The AVG play experience, once integrated with narratives, would internalize reward mechanisms within the players and in turn help foster the perception of exercise through AVG as necessary [19].

Narratives and Behavior Change Theories

The potential of using narratives can be partly explained by synergies of narrative with several widely used behavioral change theories: theory of planned behavior, social cognitive theory, and self-determination theory. Specifically, the theory of planned behavior [43] posits that a person's behavior is a function of the intention to perform that behavior, which in turn is a function of the attitude toward performing the behavior, subjective norms, and perceived behavioral control. With their immersive process, the inclusion of narratives in AVGs may make PA seem fun (changing attitudes), may show other characters engaged in PA (which may affect social norms), and can make the behavior seem easy to do (increasing perceived behavioral control), thus inducing a more positive attitude, more positive subjective norms, and greater perceived behavioral control toward performing healthy behaviors [44].

Social cognitive theory highlights observational learning, or vicarious acquisition of knowledge from the social environment, as a primary source of information [31,32]. A narrative AVG has potential through character actions to convey observational learning and useful strategies, model effective style, and demonstrate how to use these strategies, thus enhancing players' self-efficacy (the belief in one's capability to achieve different levels of performance), a key construct of social cognitive theory and a similar concept to perceived behavioral control in the theory of planned behavior. The AVG can also offer multiple vicarious experiences to the players to show the consequences of undesirable versus desirable behaviors.

Self-determination theory considers human behavior to be driven by autonomous and controlled motivation [45]. Narrative enjoyment through AVG play can be an intrinsically rewarding activity sought by people independent of extrinsic rewards by providing intriguing internal incentives for audiences who, in the role of characters, feel immersed in the story [46]. Embedding narratives into behavioral change AVGs could

potentially promote the development of autonomous motivation to complete the game and adopt the behavior promoted in the game [37].

Children's Narrative Preferences Across Media

Little is known about children's game narrative preferences. Most researchers to date have focused on print [47,48], with some later work conducted examining television [49,50]. Fiction has been recognized as the most popular choice among elementary school–aged children, with around 95% of them reading it [51]. Children have also reported that fiction is easier and more fun to read than other types of stories [52]. Most children's narrative fiction follows a linear path, with events happening in chronological order [53]. This could be partially explained by the demand to meet children's developmental needs so that they can understand the plot better.

As for the themes featured in children's fiction, various scholars have identified different methods to categorize genres. Across genders, intermediate-grade children (aged 5 to 7 years) preferred mystery and adventure stories [54], but children aged 11 to 15 years like fantasy, magic, scary, sorcery, school, romance, and true story genres [55]. Indeed, these themes help to open imaginary worlds for audiences to be transported to a new place.

Recent years have also seen a surge of discoveries about children and young adults' interest in dystopian science fictions [56-58]. While this preference may be related to the contemporary world and issues brought about by the industrial revolution and global capitalism, the thematic elements in these fictions may also appeal to children and adolescents because they offer strong emotional stimuli and allow them to explore different identities during this particular developmental stage [59].

We attempt to fill gaps in understanding regarding narrative design for AVG. Specifically, we created four different game stories based on narrative and game research with children and adolescents; 41 children between the ages of 8 and 12 years individually provided feedback on each of the four narratives and selected their favorite. We adopted a transcendental phenomenological approach, which aims to uncover the common meaning for multiple individuals of their narrative experience [60]. We then analyzed children's responses in individual cognitive interviews using an exploratory thematic analysis methodology.

Methods

Project Description

This work is part of a larger project systematically exploring the effect of narratives on children's long-term MVPA through AVG play. We created four professionally made narrative animation clips that convey information about different story arcs to be integrated seamlessly into an existing AVG that had been found to induce MVPA through the Kinect sensor on the Xbox console: Shape Up (Ubisoft) [61]. To ensure the narratives would be engaging and developmentally appropriate, we conducted individual cognitive interviews with 41 children between the ages of 8 to 12 years to gauge their interests and preferences [62].

Our key research questions include: Which of the four narratives would children consider to be the best for motivating them to exercise so that we can develop the narratives further to motivate their long-term MVPA? What are the perceived narrative characteristics that help children to exercise versus characteristics to avoid?

Sampling and Study Population

We employed a nonprobabilistic, purposive sampling approach to recruit participants from a large, diverse, urban neighborhood in the United States. Many participants came from low-income households. The 8-to-12-year age group was targeted because children younger than 8 years have cognitive limitations in responding to survey questions [63] while children older than 12 years have entered early adolescence and will be subject to many physical, mental, emotional, and social changes that may make their needs and responses different from those of younger children [64]. In addition, without intervention, obese children in this age group are highly likely to become obese young adults [65].

Since children of different weight statuses have different activity patterns when playing the selected AVG [66], we recruited approximately 20 children in each weight group (normal weight vs overweight-obese) to detect potential differences in narrative preferences. Our expectation was that 20 children in each group would be adequate to achieve theoretical saturation (ie, the point at which no new information is attained) [67].

Narrative Production

We developed four narrative plots to accompany the selected Shape Up game, which requires players to participate in a series of engaging workout exercise sessions involving kickboxing, stomping, squatting, jumping, and so on. Players can see themselves during these exercise sessions thanks to the Kinect sensor. The narrative development process was based on our previous research and experience with children's narrative engagement [25]. For example, we paid special attention to presenting and justifying the protagonist's motivation to engage in PA. Each story featured genres we believed to be appealing to children (ie, adventure, mystery, science fiction, and suspense). In accordance with the behavioral theories such as theory of planned behavior, social cognitive theory, and self-determination theory, narratives were created with the goal of encouraging children to play the selected game with increased PA intensity and duration. Each video clip lasted between 3 and 3.5 (mean 3.2) minutes, serving as a story teaser. All clips featured professional-quality animation art and voiceover with distinct styles, and all ended with a cliffhanger that dissolved into a "To be continued..." screen. More specific details about the plot, relevant theoretical concepts in narrative design, and specific design strategies for each of the stories can be found in Table 1. Overall design strategies included the following:

- We ensured that exercise and physical activities are featured throughout the four stories for observational learning while also integrating the physical activities with the narrative's natural development.

- We had the main characters engage in various types of PA and portrayed both the characters and PA as appealing and fun.

- We created character dialog and plots to elicit emotional reactions and increased enjoyment from children.

We also showed characters encountering and successfully overcoming PA challenges.

Table 1. Narrative synopses and design strategies.

Title	Synopsis	Specific design strategies
Food Fight	An adventure/quest fiction about two friends: one of them is a gamer attracted to a mobile game called "Food Fight." She accidentally spills coffee on the other friend and her phone while volunteering in a nursing home. As they are trying to dry their phones, an accidentally triggered cat-shaped timer sucks the coffee-covered friend into the world of the mobile game and turns her into a game character.	• Third-person perspective narration: female player character • Main character gender: two girls • Main character traits: helpful, considerate, brave, humorous, courageous • No antagonist • PA[a] design: character must dodge food attacks (fun and exciting) and engage in PA (jumping and running) to stay alive in game. The gamer character must figure out a way to rescue her friend from the game by engaging in all of these exercises • Reward: gamer character receives rewards as she survives each level • Modeling: characters demonstrate the exciting aspects of PA • Engagement: friendly banter and the fantastical design of the hurdles during the gamified PA acts
#PeepThisSheep	A mystery/suspense fiction about a talented child detective: the detective is lured from a party by a secret note left by someone called "9." The "9" character tries to recruit the child detective to join a top spy agency because an evil person called Cobalt plans on launching a killer app to dominate the world tomorrow evening. The agency needs the detective's help. As the detective is preparing to thwart Cobalt's app launch, a sheep video with a "PeepThisSheep" hashtag becomes wildly popular around the world.	• Second-person perspective narration: Gender-neutral player character • Main character gender: one boy and "you" • Main character traits: intelligent, warm-hearted, brave, spirited • One antagonist: Cobalt, who wants world control • PA design: detective character suddenly realizes that the video was actually the killer mind-control app and must act fast through engaging in PA (running, fighting, and searching in an interactive and simulated virtual social network) before everything gets out of control • Social support: your teammates and your classmates • Engagement: player character addresses the children directly and encourages them to exercise and demonstrates that PA is easy and fun and necessary to save the world from the evil Cobalt
Ataraxia	A fantasy/science fiction with the backdrop of a bleak future where a dictator rules the protagonist's postapocalyptic country: the character's mother adopted twin babies she found by the roadside and raised them as her own. The family later finds out that the twins do not feel pain and have the power to take pain away from others. The family tries to hide this from the world to protect the twins, but because of the twins' kind nature and natural inclination to help others, the word gets out.	• Second-person perspective narration: Gender-neutral player character • Main character gender: a boy and a girl and "you" • Main character traits: courageous, nice, compassionate, adventurous • One antagonist: evil dictator • PA design: the evil dictator discovers this and abducts the twins so he can use their genes to create a force of invincible super soldiers; the character must stop him through various PA engagement (searching, jumping, running, and combat on a future planet) • Modeling: the twins keep engaging in fun and exciting PA • Social support: the mother figure helps the twins and you to overcome hurdles for PA • Engagement: emotional connection with the twins as your "siblings" through vivid details
Star Dust	A mystery/suspense fiction: it begins with two friends' over-night field trip to an observatory to watch a meteor shower with their classmates and parents, where an eccentric professor greets them. The protagonist discovers an ancient prophecy while inside the observatory: the meteor shower will bring alien matter and wipe out the human race. When the shower begins, everyone outside starts to behave erratically.	• Third-person perspective narration: male player character • Main character gender: a boy and a girl • Main character traits: resourceful, fearless, smart, quick-witted • One antagonist: a scientist who keeps the remedy to the alien matter away from everyone • PA design: character must figure out how to deal with those who are affected and save the world through engaging in various PA behaviors (dodging infected people, investigating the environment, and searching for a remedy to the alien matter) • Engagement: player character needs to solve many puzzles to find out how to save the best friend who may have been infected by the scientist by overcoming many hurdles on the quest

[a]PA: physical activity.

Data Collection and Analysis

Ethical approval was obtained from the Northeastern University institutional review board. We obtained written informed consent (from parents or guardians) and assent (from children) from all participants. Data were collected between October 2016 and March 2017 at a local community center for families.

XSL·FO
RenderX

Demographic information was collected from parents or guardians.

Three research assistants included undergraduate and graduate students majoring in psychology, public health, and communication. Each of them received 25 to 30 hours of extensive training in cognitive interview techniques prior to data collection. They used an interview protocol consisting of 5 structured demographic questions, followed by 8 open-ended cognitive interview questions, and four open-ended follow-up questions for each narrative. Interview questions were as follows:

- Tell me what happened in the story?
- What do you think happens next?
- How do you think the story will end?
- What did you think about the story?
- Was there anything about the story that was hard to understand?
- If you were the writer, would you make any changes to the story?
- Now let's talk about the characters. What did you think about the characters?

- Which of these characters might be able to help you move around and be more active?

Follow-up questions were asked only if a participant's response to the initial question did not cover specific topics of interest.

The four narratives were shown to each child in a random sequence on a laptop to minimize order effects. After each child viewed each of the four narratives and completed the cognitive interviews, the assistants gave them four photo cards (Figure 1) showing the title and a screenshot of each narrative and asked each participant to rank the narratives from most favorite to least favorite by arranging the cards on a table. The assistants documented each child's narrative rankings (1=most favorite; 4=least favorite) and then asked them the reasons for their rankings. Children were allowed to take a 5-minute break between each story interview, although few did due to their enthusiasm about the animated narratives. Sessions lasted between 50 and 64 minutes. All children answered all questions. Each child received a $25 gift card and they were entered into a drawing to win an OgoDisk set (OgoSport LLC).

Figure 1. Picture Cards for Narratives.

Food Fight

#PeepThisSheep

Ataraxia

Star Dust

The inclusion criteria were between ages 8 and 12 years; able to speak English; and no visual, intellectual, or neurodevelopmental conditions (eg, autism, anxiety disorders).

All interviews were conducted in English and digitally recorded. Verbatim responses were transcribed by a professional transcription agency. We then compared the transcripts to

recordings for accuracy and revised the transcripts when necessary. To minimize the likelihood of unintentional bias, three coders (undergraduate and graduate students majoring in public health and communication) independently analyzed the transcripts [68]. Coders received around 35 to 40 hours of training for coding qualitative data and cognitive interviews as part of the process of sensitizing concepts [69]. They read each of the 41 transcripts multiple times before the coding process began.

Coding used a hybrid thematic analytic approach. A structured approach was initially used with interview questions providing the framework. Within each question, similar answers were grouped and assigned an emergent code (eg, positive characteristics of a character) [70]. We organized and grouped the codes for each question into higher order codes and then into themes using an open-coding process. The themes were arranged into two categories: features that motivate children to exercise versus features that discourage them from exercising. This strategy was applied to transcripts for all narratives. The coding team met weekly over 3 months with the first author to discuss the codes and organize them into themes within each category. We achieved a general consensus for each category over the meetings and observed theoretical saturation across the interviews.

Results

Participant Characteristics

The two weight groups were comparable in multiple characteristics (Table 2). Children of both groups had a moderate degree of experience in playing Kinect games.

Table 2. Children's demographic information (N=41).

Characteristic	Normal weight n=21	Overweight-obese n=20	*P* value
Age in years, mean (SD)	10 (1.6)	10.9 (1.4)	.51
Gender	__[b]	—	**.26**
Boy	10	13	—
Girl	11	7	—
BMI[a], kg/m^2 (SD)	17 (1.9)	25.8 (3.3)	.01
BMI, percentile (SD)	49.7 (26.7)	96.3 (2.1)	.01
Race, n (%)	—	—	**.67**
African American	10 (48)	7 (35)	—
American Indian/Alaska Native	2 (10)	1 (5)	—
Asian	1 (5)	2 (10)	—
Caucasian	5 (24)	5 (25)	—
Other (mixed)	3 (14)	5 (25)	—
Parent education, n (%)	—	—	**.31**
High school	5 (24)	1 (5)	—
Technical school	1 (5)	2 (10)	—
Some college	2 (10)	4 (20)	—
College graduate	4 (19)	8 (40)	—
Postgraduate Study	9 (43)	5 (25)	—
Annual household income (US $)[c], n (%)	—	—	**.76**
<$20,000	6 (29)	4 (21)	—
$20,000-$39,999	5 (24)	7 (37)	—
$40,000-$59,999	1 (5)	1 (5)	—
$60,000-$79,999	3 (14)	3 (16)	—
$80,000-$99,999	4 (19)	1 (5)	—
>$100,000	2 (10)	3 (16)	—
Kinect games (KG) experience	—	—	—
How much have you played KG? (1=a little; 5=a l lot)	2 (1.5)	2.6 (1.3)	.15
How familiar are you with KG? (1=unfamiliar; 5=familiar)	2.3 (1.4)	2.7 (1.6)	.61

[a]BMI: body mass index.

[b]Not Applicable

[c]One family of an overweight-obese child did not provide an answer to this question. Thus, n=19.

Narrative Preferences

We calculated the preference score distribution and average rank score for each narrative. A lower score indicates higher preference (eg, the favorite story was ranked 1 most favorite, second favorite as 2, and so on). Ataraxia was the top-rated story overall and among all weight, gender, and racial groups. The story takes place in a dystopian future, when a twin brother and sister with the ability to absorb and take away people's pain have been kidnapped by an evil dictator to create an army of indestructible soldiers. In addition, the interview transcripts indicated children's negative comments and criticisms were the shortest for Ataraxia compared to the other three options. In general, we did not observe differences across the weight, gender, and racial groups in responses to the narratives; therefore, the ratings of each story were collapsed across all groups (Table 3).

Table 3. Children's narrative preferences.

Title	1 (most favorite)	2	3	4 (least favorite)	Average
Ataraxia	14	16	9	2	1.98
Food Fight	11	5	15	10	2.59
Star Dust	7	11	13	10	2.63
#PeepThisSheep	9	9	4	19	2.80

Thematic Analyses

The cognitive interview analyses provided rich insights regarding key characteristics of stories (eg, characters and plot) children found effective versus ineffective for motivating them to exercise. Analyses included comments regarding children's evaluative responses (cognitive and affective) to other narrative characteristics related to our research questions. In comparing feedback, few differences were observed among children of both weight groups. Therefore, their responses were combined. To maintain confidentiality, participant quotes are identified with a randomly generated number between 0 and 300.

Responses to the Favorite Narrative

Ataraxia, a science fiction story set in the dystopian future, was selected as the favorite narrative among all weight, gender, and race groups. Most of the themes identified to motivate children to exercise can be found from children's response to this story. Therefore, they are mentioned here in the order of the frequency observed with a special emphasis on Ataraxia. More specifically, the themes revolved around protagonists, character attributes, and story elements in the order of their observed weight. In the next section, after illustrating the themes with quotes, we connected findings with relevant concepts from health behavior and media psychological theories.

Almost all children identified the twins (ie, the protagonists) as appealing protagonists who would motivate them to exercise. The primary reason usually mentioned was that the activities the twins were performing (eg, leaping from roofs, running after wild boars, or playing with nests of hornets) were fun and exciting.

> *I think I might want to join them because it looked a little bit fun.* [Child 83]

The children seemed to have a positive attitude toward PA, which is a crucial determinant of the PA behavioral intention, the strongest predictor of PA behavior according to the theory of planned behavior. This also echoes social cognitive theory's emphasis on observational learning via role models and self-determination theory's emphasis on enjoyment in intrinsically motivating children to engage in PA.

The second theme was that the characters had good moral qualities, which echoes the findings that audiences tend to engage with moral characters [71].

> *Both of them were really brave and really courageous because not only they stayed in the house, but they jumped over...from house to house.* [Child 217]

> *They could probably teach me how to be more active and teach me how to be generous that they were in the story.* [Child 67]

To a certain extent, likeable characters also make social cognitive theory's observational learning more likely to happen when children identify with the credible characters perceived to have good moral qualities. Similarly, according to self-determination theory, which specifies the basic psychological needs for relatedness to be one of the key precursors for autonomous motivation and engagement, characters perceived to have good moral qualities would foster relatedness by bring children closer to the character who engages in PA. Additionally, such qualities would help to enhance children's emotional engagement and connections with the characters, thus potentially enhancing immersion.

Last, some overweight-obese children and a few normal weight children mentioned that the twins have special abilities that enable them not to feel pain, which is desirable for exercise.

> *...so if you get hurt then you can still run around.* [Child 163]

This is a notable example of narrative's power to enhance perceived behavioral control, another key influencer of the behavioral intention, which determines behavior according to theory of planned behavior. Enhanced perceived behavioral control could also potentially increase children's self-efficacy, or confidence in performing a particular behavior according to social cognitive theory. The self-efficacy increase could occur through observational learning or modeling of characters. In addition, self-efficacy can also be enhanced through character identification, when children put themselves into the character's shoes. This could help children overcome fear of discomfort in performing PA and increase their confidence when they start identifying with the twin characters. Interestingly, quite a few responses suggested strong parasocial interaction [72] with the twins.

> *...If I was playing with them...like hanging out with them...* [Child 97]

> *...If they were my friends, they'd be like cool people to hang out with.* [Child 292]

These coincided with the positive role model effect and potential emotional affinity and engagement (ie, connection) with the characters. According to self-determination theory, these are important factors helping children to develop autonomous motivation for PA behaviors.

Aside from the twins, other characters also played a role in motivating the participants to exercise. For example, the mother

"...seems very nice and she seems like she could give me some confidence" [Child 211].

It seems that the mother character had a positive effect on the subjective norms (ie, children's beliefs about how people they care about would view their PA behavior according to theory of planned behavior). The mother character's encouraging gestures and actions would also potentially increase the children's self-efficacy in PA engagement. The character "I," who was not shown in the narrative directly, could enhance PA motivation as evidenced by "I get jealous of the video game characters, that they get to do all this running around" [Child 212] or because I want to be "try[ing] to get [my] [family] back" [Child 161], all of which suggested potential emotional engagement.

Interestingly, opinion about the antagonist (evil dictator) was split. While some considered him as a character who would help them to exercise, the action was more of "run[ning] away from him" [Child 158] than running along with him. Many others did not consider an evil character to be trustworthy or motivational.

> *If this guy is secretly a super villain who wants some mindless soldiers, I don't think a lot of people would be wanting to do what he does.* [Child 2]

These responses suggested the importance of positive and engaging narrative character creation, as character with negative qualities may not be as effective at motivating children to perform the desired PA behaviors.

Another theme identified was the story's immersive characteristics.

> *You can join the story... That would be cool...[and] you are part of the story.* [Child 5]

> *I liked how you were so...like you in the story.* [Child 8]

> *I was really intrigued and everything. When it said they [the twins] were gone, I was just like, "Don't end, don't end, don't end," and then it ended.* [Child 44]

Many children used the words "interesting," "cool," "fun," and "creative" to describe their overall impression of the narrative. Children also appreciated the amount of detail and explanations offered throughout the plotline.

> *I think that's a good thing because there was lots of like plot into the story and lots of details, so I think that was good.* [Child 103]

> *[Be]cause it gave reasons for everything...because it explained everything.* [Child 162]

They also expressed their desire to continue following the story development.

> *Yeah. I like things that will continue in series. It'll keep on going and going and going until like the very last.* [Child 194]

Both children's affinity for detail and need for additional story development also suggested their high level of narrative involvement with Ataraxia [73], which could potentially lead to increased PA later.

Responses to the Remaining Narratives

Additional themes were identified from the other three narratives to explain why the children did not find them as effective exercise motivators. While in general children liked the art styles and plots of these narratives, many used "boring" to describe the characters and plots in the remaining narratives. Some meaningful patterns emerged when we further examined their comments. The results are also focused on protagonists, character attributes, and story elements.

First, there was a lack of interesting or exciting action, which might make the protagonist less attractive. For example, when talking about Food Fight, one child commented, "I think it's boring because, well the character didn't actually fight in the game" [Child 5]. Another said that "It's like—it needs more action" [Child 239]. The same thought was mentioned for #PeepThisSheep as well, "[Be]cause it doesn't have that much action" [Child 271]. Similarly, another child mentioned that the reason Star Dust was not ranked higher because "...[w]alking, running. It's all they do" [Child 193]. From a theory of planned behavior framework, these narratives did not seem to help with the participants' subjective norms by showing PA behaviors as part of the desired action constantly performed by the characters. The narratives failed to provide accessible and attractive behavioral models for children to learn through observation of the character actions according to social cognitive theory. These issues may lead to the failure to instigate autonomous motivation from children, let alone enjoyment, both of which are important in instigating PA behaviors according to self-determination theory.

Second, lack of details about characters and plots seemed to have prevented children from fully understanding and appreciating the narrative developments. For example, when talking about #PeepThisSheep, a child said that the narrative "did not have that much details in it...for [the child] to understand what that means" [Child 216]. Another mentioned adding a bit information to the antagonist.

> *Maybe just [like] give a little back story on why she wants to release this Peep this Sheep on social media.* [Child 2]

Similarly, another child suggested offering additional information for how the character in Food Fight was stuck in the phone.

> *They could give a whole explanation of how she got stuck in the phone.* [Child 162]

While the exact causal relationship between details and narrative engagement needs additional exploration from an interdisciplinary angle, children's complaints about the lack of narrative detail suggested their lower level of character appreciation and narrative engagement, which could be less likely translated into PA motivation.

Third, the narrative motifs were perceived to rely on common or overused literary tropes, which might make the plot less interesting. For example, when one child described Star Dust, they said:

> *Because it was boring. How like—they just go on a camping trip. And then everything just goes bad. And like everybody just starts turning evil. I feel like that's what happens in most stories.* [Child 156]

Another described Food Fight by saying:

> *I'm not saying that Food Fight is bad or anything; it's just that it could have...been a little bit more creative.* [Child 2]

Similarly, when describing Cobalt, the antagonist from #PeepThisSheep, one child said:

> *Cobalt seems like a character I've seen before inside a movie.* [Child 96]

These responses suggested repeated narrative tropes may not bring about the optimal engagement from children to attend to the story development, let alone participate in PA.

In addition, children seemed to take it for granted that there should be a bad guy (antagonist) in the story.

> *I like that there's a video and everybody likes to watch it, so a bad guy could hatch a plan.* [Child 163, #PeepThisSheep]

> *Cause he's a scientist and knows about like a lot of stuff and then he's so evil, but I think evil's pretty cool.* [Child 96, Star Dust]

> *Because like if a story doesn't have a bad guy, then it's just going to be boring.* [Child 156, Ataraxia]

On the other hand, as Food Fight did not feature a true antagonist and shifted the conflicts among the two friends, it did not seem to work for motivating children to exercise.

> *I have a best friend and me and her, we don't get into like fights a lot because we know it was gonna ruin our friendship together* [Child 83]

> *Because it's better to say sorry than just get mad at your friends.* [Child 173]

These responses emphasize the importance of creating interesting antagonists with depth and dimension to maintain children's interest.

Last, children did not particularly enjoy the spy-themed plot in #PeepThisSheep. More specifically, they seemed to be bothered by how the main character was approached and recruited by "9," who lured the main character out of a school party with a favorite snack and left a secret note. Quite a few children seemed to be irritated by his actions.

> *"9" was so secretive; he was kind of creepy.* [Child 8]

Others equated his behavior to that of the antagonist.

> *Well, [what] I don't get the plan is why "9" would be stalking you instead of [Cobalt].* [Child 161]

Accordingly, they questioned the subsequent plot development.

> *It was kind of hard to understand how "9" is a stranger to the narrator and then like [the narrator] just ends up going to his basement and...joining...his spy team.* [Child 167]

Discussion

Practical Recomendations

Based on the exploratory thematic analysis, we arranged the two themes identified (features that motivate children to exercise vs features that discourage them to exercise) into preliminary creative design principles. The principles are presented as Dos and Don'ts for narrative creation intended to motivate children to participate in physical activities. The theoretical implications are also discussed. Please note that given the limited sample size and research strategy, all of the recommendations are presented more as exploratory possibilities that need to be empirically verified.

Dos

Do Create Child Characters With Strong Moral Values

Children identify with and want to model themselves after characters who are good people and who are like them. Both tendencies have been well documented by social learning theory and affective disposition theory researchers [74,75]. While truly bad characters are also needed as antagonists, they may motivate children to be physically active but in a more passive way. For example, the children can escape when the bad guys are threatening their lives. On the other hand, children do not want to see the protagonists fight against each other.

Do Involve Extraordinary Actions

Designers should create narratives with characters in constant need of action throughout, or in other words, action-packed narratives. The actions should not be regular mundane routines (eg, walking or running) but something unusual. This concept is closely related to the interactive nature of active games and will also cater well to children's novelty-seeking tendency in their developmental stage to capture and sustain their attention to the story development.

Do Make the Desired Exercise Fun, Interesting, and Integrated Into the Story

Excitement is crucial for inciting children to imitate characters' actions and ultimately initiate their own PA. Fun and interesting exercise makes PA more desirable and less tedious [76]. When excitement, fun, and interesting exercise are packaged within an active game, children will more likely be intrinsically motivated to exercise over time [77].

Do Involve Super Powers When Creating Protagonists and Plots

Fantasies and fictions open up imaginary worlds, enabling children to be transported to the narrative world [78]. Super powers not only could make the physical activities engaging

XSL•FO

RenderX

and exciting but also may potentially help remove perceived obstacles for children who are not used to exercise and who would not want to exercise, to increase their self-efficacy.

Do Create Immersive Story Plots With Intriguing Cliffhangers

The power of the immersive narrative engagement has been discussed extensively [79]. In addition, suspense is a frequently employed device for maintaining audience interest in stories [80]. Intriguing cliff hangers delivered at the right moment may make children curious about what happens next [81]. They will be eager to continue their narrative engagement through additional exercise behaviors.

Don'ts

Don't Omit Information From the Story Development

To ensure that the narratives are organically related to physical activities, relevant information about why the character is engaging in PA should be integrated into the story development. This will help to justify character's motivation to engage in different types of exercises as a natural progression of plot development, thus making character actions believable and improving the players' intrinsic motivation to participate in physical activities.

Don't Forget to Provide Reasonable Explanations (With Details) for Why Things Happen

While adults may make the mental bridges as part of their sense-making process to understand and process narratives, the extra effort required may be taxing for children [82] and may reduce their narrative engagement. It is important to assess if a narrative has a well-balanced portion of information given and withheld from children so that their interest in the characters and plot can be constantly piqued but not burdened with too many questions along the way.

Don't Put Protagonists in a Passive Situation to Be Tricked or Watched

An important feature of interactive games is that the players are allowed to be active agents making interactive exchanges within the game to make progress [83]. As a result, while it is reasonable to have some obstacles for the protagonists during the narrative, they should still be given a moderate amount of agency to be in charge of their fate. Putting them in a passive situation (eg, being stalked or watched) may reduce children's motivation to engage in PA.

Don't Make Boring Stories With a Lot of Ordinary Events or Actions

Children are savvy media consumers [84]. Narrative creators should not underestimate their expectation of innovative stories with unpredictable endings. In addition, they want their imagination to be their guide as they are transported into a fantastical new world instead of getting trapped in the ordinary daily routines.

Limitations

A limitation of this study is that we relied exclusively on self-report interview responses instead of measuring children's actual PA behavior in real-world settings. Further studies are needed to discover if narratives produced with these features in mind would indeed lead children to actually play more AVGs. Additionally, the long-term effect of the narrative's motivation should be validated by children's play behavior over time. We only used one story from each genre, and it may be hard to pinpoint why exactly children preferred one story over the others due to the different characteristics of the four stories. The animated narratives were not integrated into the game, Shape Up, during this round of data collection, and children did not have the opportunity to play the games after they viewed the videos. Additional studies are needed to explore whether integration of Ataraxia into Shape Up gameplay would motivate children to exercise. Another limitation is that we have a relatively small sample size and only conducted this study in one area of the country; recruited children across many weight, gender, and racial groups from low-income households; and focused primarily on game narrative development, all of which would limit the generalizability of our findings. In other words, while we did not observe differences across the groups, we want to add the caveat that the subgroups were relatively small and could limit our capacity to detect group difference. All of participants, however, seemed eager to participate in such projects and were able to provide insightful responses. Last, we only had one narrative for each genre, which could serve as a confounder. It is difficult to tease apart whether the design strategy or the theme of each narrative was the actual reason for children's preference. It would be ideal if two or more variations of the narratives within the same theme could be designed and tested with children [85].

Conclusion

Narratives have immense potential for promoting PA among children through AVGs, but less is known about active game narrative development. Our work here serves as one of the first steps toward a series of systematic inquiries into how to maximize the behavioral potential of narratives for combating childhood obesity. We tried to incorporate existing interdisciplinary psychological and behavioral theories into the narrative development and used direct feedback from our participants to further inform our narrative development process.

A clinically relevant age group (8 to 12 years) was chosen for this project because they are most at risk and closely related to the obesity issue and the age group would be a proper range to intervene. Given the unique persuasive power of narratives, it is also likely that a story developed with this age group in mind would be mostly likely influential among the children. We believe this kind of process will help to create an iterative design methodology for the development of child-friendly narratives that are effective in motivating PA in the future.

From a theoretical perspective, we tried to integrate multiple concepts from behavioral theories (theory of planned behavior, social cognitive theory, self-determination theory, etc) and media psychology (identification, interpersonal attraction, parasocial interaction, etc) into the narrative development stage (see Table 1 for details) and later explored their engagement effect through a close analysis of children's feedback in the result section. Children seemed to be attracted to positive role models who are engaged in unique forms of PA. Such role

XSL•FO

RenderX

models increased their narrative involvement and, to some extent, their self-efficacy. Children also highlighted the way the narratives created autonomous motivation (eg, wanting to be like the characters or spend time with the characters) and increased their subjective norms (eg, important characters encouraging them to exercise).

This research provides preliminary evidence that narratives have characteristics which may increase child PA when playing active games. Appealing features include positive characters, extraordinary actions, interesting plots, super powers, as well as engaging cliffhangers, all of which should be aligned with the exercise motivation. On the other hand, mentally taxing, passive protagonists devoid of agency and hackneyed narrative tropes should be avoided. Future work is needed to verify these findings and examine their effect of PA during gameplay.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Warren M, Beck S, Rayburn J. Trust for America's Health (TFAH) and Robert Wood Johnson Foundation. 2018. State of obesity: better policies for a healthier America URL: https://www.tfah.org/wp-content/uploads/2018/09/TFAH-2018-ObesityReport-FINAL.pdf [accessed 2019-10-19]
2. World Health Organization. 2018. Obesity and overweight URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2019-07-26]
3. Skinner AC, Perrin EM, Skelton JA. Prevalence of obesity and severe obesity in US children, 1999-2014. Obesity (Silver Spring) 2016 May;24(5):1116-1123 [FREE Full text] [doi: 10.1002/oby.21497] [Medline: 27112068]
4. National Institute of Diabetes and Digestive and Kidney Diseases. 2018. Health risks of overweight & obesity URL: https://www.niddk.nih.gov/health-information/weight-management/adult-overweight-obesity/health-risks [accessed 2019-07-26]
5. Janssen I, Katzmarzyk PT, Boyce WF, Vereecken C, Mulvihill C, Roberts C, Health Behaviour in School-Aged Children Obesity Working Group. Comparison of overweight and obesity prevalence in school-aged youth from 34 countries and their relationships with physical activity and dietary patterns. Obes Rev 2005 May;6(2):123-132. [doi: 10.1111/j.1467-789X.2005.00176.x] [Medline: 15836463]
6. 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee scientific report. Washington: US Department of Health and Human Services; 2018.
7. National Physical Activity Plan Alliance. 2016. The 2016 United States report card on physical activity for children and youth URL: http://www.physicalactivityplan.org/reportcard/2016FINAL_USReportCard.pdf [accessed 2018-08-21]
8. Metcalf B, Henley W, Wilkin T. Republished research: effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). Br J Sports Med 2013 Mar;47(4):226. [doi: 10.1136/bjsports-2013-e5888rep] [Medline: 23401246]
9. Bailey BW, McInnis K. Energy cost of exergaming: a comparison of the energy cost of 6 forms of exergaming. Arch Pediatr Adolesc Med 2011 Jul;165(7):597-602. [doi: 10.1001/archpediatrics.2011.15] [Medline: 21383255]
10. Gao Z, Chen S, Pasco D, Pope Z. A meta-analysis of active video games on health outcomes among children and adolescents. Obes Rev 2015 Sep;16(9):783-794. [doi: 10.1111/obr.12287] [Medline: 25943852]
11. Entertainment Software Association. 2019. Essential facts about the computer and video game industry URL: https://www.theesa.com/wp-content/uploads/2019/05/ESA_Essential_facts_2019_final.pdf [accessed 2019-07-26]
12. Gaming Weekender. 2019 Jun 11. Best game console for exercise in 2019? (Switch vs. Xbox One vs. PS4) URL: https://www.gamingweekender.com/best-game-console-for-exercise/ [accessed 2019-07-26]
13. Tappe KA, Glanz K, Sallis JF, Zhou C, Saelens BE. Children's physical activity and parents' perception of the neighborhood environment: neighborhood impact on kids study. Int J Behav Nutr Phys Act 2013 Mar 27;10:39 [FREE Full text] [doi: 10.1186/1479-5868-10-39] [Medline: 23531282]
14. Lyons EJ, Tate DF, Ward DS, Bowling JM, Ribisl KM, Kalyararaman S. Energy expenditure and enjoyment during video game play: differences by game type. Med Sci Sports Exerc 2011 Oct;43(10):1987-1993 [FREE Full text] [doi: 10.1249/MSS.0b013e318216ebf3] [Medline: 21364477]
15. Abernathy T, Rouse R. Death to the three-act structure: toward a unique structure for game narratives. 2014. URL: http://www.gdcvault.com/play/1020050/Death-to-the-Three-Act [accessed 2014-07-16]
16. Chin A Paw MJM, Jacobs WM, Vaessen EPG, Titze S, van Mechelen W. The motivation of children to play an active video game. J Sci Med Sport 2008 Apr;11(2):163-166. [doi: 10.1016/j.jsams.2007.06.001] [Medline: 17706461]

XSL•FO

RenderX

17. Graves LEF, Ridgers ND, Atkinson G, Stratton G. The effect of active video gaming on children's physical activity, behavior preferences and body composition. Pediatr Exerc Sci 2010 Nov;22(4):535-546. [Medline: 21242603]

18. Epstein LH, Roemmich JN. Reducing sedentary behavior: role in modifying physical activity. Exerc Sport Sci Rev 2001 Jul;29(3):103-108. [Medline: 11474956]

19. Lu AS, Baranowski T, Thompson D, Buday R. Story immersion of videogames for youth health promotion: a review of literature. Games Health J 2012 Jun;1(3):199-204 [FREE Full text] [doi: 10.1089/g4h.2011.0012] [Medline: 24416639]

20. Lu AS, Kharrazi H. A state-of-the-art systematic content analysis of games for health. Games Health J 2018 Feb;7(1):1-15 [FREE Full text] [doi: 10.1089/g4h.2017.0095] [Medline: 29293368]

21. Lu AS, Kharrazi H, Gharghabi F, Thompson D. A systematic review of health videogames on childhood obesity prevention and intervention. Games Health J 2013 Jun;2(3):131-141 [FREE Full text] [doi: 10.1089/g4h.2013.0025] [Medline: 24353906]

22. Baranowski MT, Lu AS, Buday R, Lyons EJ, Schell J, Russoniello C. Stories in games for health: more pros or cons? Games Health J 2013 Oct;2(5):256-263 [FREE Full text] [doi: 10.1089/g4h.2013.0052] [Medline: 26196925]

23. Lu AS, Baranowski T, Hong SL, Buday R, Thompson D, Beltran A, et al. The narrative impact of active video games on physical activity among children: a feasibility study. J Med Internet Res 2016 Oct 14;18(10):e272 [FREE Full text] [doi: 10.2196/jmir.6538] [Medline: 27742605]

24. Hwang J, Lu AS. Narrative and active video game in separate and additive effects of physical activity and cognitive function among young adults. Sci Rep 2018 Jul 20;8(1):11020 [FREE Full text] [doi: 10.1038/s41598-018-29274-0] [Medline: 30030456]

25. Lu A, Buday R, Thompson D, Baranowski T. What type of narrative do children prefer in active video games? An exploratory study of cognitive and emotional responses. In: Tettegah S, Huang WH, editors. Emotions, Technology, and Digital Games. London: Elsevier Publications; 2016.

26. Fernandes Davies V, Mafra R, Beltran A, Baranowski T, Lu AS. Children's cognitive and affective responses about a narrative versus a non-narrative cartoon designed for an active videogame. Games Health J 2016 Apr;5(2):114-119 [FREE Full text] [doi: 10.1089/g4h.2015.0019] [Medline: 26881473]

27. Lu AS. Narrative in exergames: thoughts on procedure, mechanism, and others. Games Health J 2015 Feb;4(1):19-24 [FREE Full text] [doi: 10.1089/g4h.2014.0090] [Medline: 26181676]

28. Rimmon-Kenan S. Narrative Fiction: Contemporary Poetics. 2nd Edition. London: Routledge; 2002.

29. Cohen J. Defining identification: a theoretical look at the identification of audiences with media characters. Mass Comm Soc 2001 Aug;4(3):245-264. [doi: 10.1207/S15327825MCS0403_01]

30. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000 Nov;79(5):701-721. [doi: 10.1037//0022-3514.79.5.701] [Medline: 11079236]

31. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]

32. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]

33. Schultz J. Language in the inner city: studies in the black English vernacular. In: Labov W, editor. Urban Life. Philadelphia: University of Philadelphia Press; Sep 15, 2016:490-492.

34. van Laer T, de Ruyter K, Visconti LM, Wetzels M. The extended transportation-imagery model: a meta-analysis of the antecedents and consequences of consumers' narrative transportation. J Consum Res 2014 Feb 01;40(5):797-817. [doi: 10.1086/673383]

35. Murphy ST, Frank LB, Chatterjee JS, Moran MB, Zhao N, Amezola de Herrera P, et al. Comparing the relative efficacy of narrative vs nonnarrative health messages in reducing health disparities using a randomized trial. Am J Public Health 2015 Oct;105(10):2117-2123. [doi: 10.2105/AJPH.2014.302332] [Medline: 25905845]

36. Oatley K. Emotions and the story worlds of fiction. In: Green MC, Strange JJ, Brock TC, editors. Narrative Impact: Social and Cognitive Foundations. Mahwah: Lawrence Erlbaum Associates; 2002:39-69.

37. Pillay H. An investigation of cognitive processes engaged in by recreational computer game players. J Res Technol Educ 2002 Mar;34(3):336-350. [doi: 10.1080/15391523.2002.10782354]

38. Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. Educ Psychol 2003 Mar;38(1):43-52. [doi: 10.1207/S15326985EP3801_6]

39. Forrester MA. Can narratology facilitate successful communication in hypermedia environments? Intelligent Tutoring Media 1996 Jul;7(1):11-20. [doi: 10.1080/14626269609408364]

40. Slater MD, Rouner D. Entertainment, education and elaboration likelihood: understanding the processing of narrative persuasion. In: Green MC, Strange JJ, Brock TC, editors. Narrative Impact: Social and Cognitive Foundations. Mahwah: Lawrence Erlbaum Associates; May 2002:173-191.

41. Hefner D, Klimmt C, Vorderer P. Identification with the player character as determinant of video game enjoyment. In: Entertainment Computing ICEC. Berlin: Springer; 2007.

42. McLellan H. Hypertextual talestory models for hypertext design. J Educ Multimedia Hypermedia 1993;2(3):239-260.

43. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]

44. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]

45. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]

46. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]

47. Thorndike R. A Comparative Study of Children's Reading Interests. New York: Columbia University Press; 1941.

48. Chapman M, Filipenko M, McTavish M, Shapiro J. First graders' preferences for narrative and/or information books and perceptions of other boys' and girls' book preferences. Can J Educ 2007;30(2):531-553. [doi: 10.2307/20466649]

49. Valkenburg P, Janssen S. What do children value in entertainment programs? A cross-cultural investigation. J Comm 1999;49(2):3-21. [doi: 10.1111/j.1460-2466.1999.tb02790.x]

50. Bryant J, Anderson D. Children's Understanding of Television: Research on Attention and Comprehension. Cambridge: Academic Press; 1983.

51. Clark C, Rumbold K. National Literacy Trust. 2006. Reading for pleasure: a research overview URL: https://literacytrust.org.uk/research-services/research-reports/reading-pleasure-research-overview/ [accessed 2019-05-01]

52. Moyer J. Learning from leisure reading. Ref User Serv Q 2007;46(4):66-79. [doi: 10.5860/rusq.46n4.66]

53. Horning KT. From Cover to Cover: Evaluating and Reviewing Children's Books. New York: Harper Collins; 2010.

54. Summers EG, Lukasevich A. Reading preferences of intermediate-grade children in relation to sex, community, and maturation (grade level): a Canadian perspective. Reading Res Q 1983;18(3):347-360. [doi: 10.2307/747393]

55. Hopper R. What are teenagers reading? Adolescent fiction reading habits and reading choices. Literacy 2005 Nov;39(3):113-120. [doi: 10.1111/j.1467-9345.2005.00409.x]

56. Sambell K. Presenting the case for social change: the creative dilemma of dystopian writing for children. In: Hintz P, Ostry E, editors. Utopian and Dystopian Writing for Children and Young Adults. London: Routledge; 2013:175-190.

57. Flanagan V. Technology and Identity in Young Adult Fiction: The Posthuman Subject. New York: Springer; 2014.

58. Basu B, Broad K, Hintz C. Contemporary Dystopian Fiction for Young Adults: Brave New Teenagers. London: Routledge; 2013.

59. Steinberg L. Age of Opportunity: Lessons From the New Science of Adolescence. New York: Houghton Mifflin Harcourt; 2014.

60. Creswell J, Poth C. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. 4th Edition. Thousand Oaks: Sage Publishing House; 2017.

61. Hwang J, Fernandez AM, Lu AS. Application and validation of activity monitors' epoch lengths and placement sites for physical activity assessment in exergaming. J Clin Med 2018 Sep 11;7(9) [FREE Full text] [doi: 10.3390/jcm7090268] [Medline: 30208567]

62. Willis G. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks: Sage Publishing House; 2005.

63. Borgers N, de Leeuw E, Hox J. Children as respondents in survey research: cognitive development and response quality. Bull Sociol Methodol 2016 Jul 22;66(1):60-75. [doi: 10.1177/075910630006600106]

64. Centers for Disease Control and Prevention. 2010. Child development URL: http://www.cdc.gov/ncbddd/child/earlyadolescence.htm [accessed 2011-02-07]

65. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. N Engl J Med 1997 Sep 25;337(13):869-873. [doi: 10.1056/NEJM199709253371301] [Medline: 9302300]

66. Hwang J, Lee I, Fernandez AM, Hillman CH, Lu AS. Exploring energy expenditure and body movement of exergaming in children of different weight status. Pediatr Exerc Sci 2019 Aug 08:1-10. [doi: 10.1123/pes.2019-0006] [Medline: 31394495]

67. Kreuger R, Casey M. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks: Sage Publications; 2009.

68. Barbour R. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? BMJ 2001 May 05;322(7294):1115-1117 [FREE Full text] [doi: 10.1136/bmj.322.7294.1115] [Medline: 11337448]

69. Miles M, Huberman A, Saldana J. Qualitative Data Analysis. Thousand Oaks: Sage Publications; 2014.

70. Walker D, Myrick F. Grounded theory: an exploration of process and procedure. Qual Health Res 2006 Apr;16(4):547-559. [doi: 10.1177/1049732305285972] [Medline: 16513996]

71. Zillmann D. The psychology of suspense in dramatic exposition. In: Vorderer P, Wulff HJ, Friedrichsen M, editors. Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations. Hillsdale: Lawrence Erlbaum Associates, Inc; 1996:199-231.

72. Hoffner C. Children's wishful identification and parasocial interaction with favorite television characters. Journal of Broadcasting & Electronic Media 1996 Jun;40(3):389-402. [doi: 10.1080/08838159609364360]

73.   Tal-Or N, Cohen J. Understanding audience involvement: Conceptualizing and manipulating identification and transportation. Poetics 2010 Aug;38(4):402-418. [doi: 10.1016/j.poetic.2010.05.004]

74.   Raney AA. The psychology of disposition-based theories of media enjoyment. In: Bryant J, Vorderer P, editors. Psychology of Entertainment. Mahwah: Lawrence Erlbaum Associates Publishers; 2006:137-150.

75.   Rottschaefer W. Social learning theories of moral agency. Behav Philosophy 1991;19(1):61-76.

76.   Garn A, Cothran D. The fun factor in physical education. J Teach Phys Educ 2006;25(3):281-297. [doi: 10.1123/jtpe.25.3.281]

77.   Ferrer-Caja E, Weiss MR. Predictors of intrinsic motivation among adolescent students in physical education. Res Q Exerc Sport 2000 Sep;71(3):267-279. [doi: 10.1080/02701367.2000.10608907] [Medline: 10999264]

78.   Green M, Brock T. In the Mind's Eye: Transportation-Imagery Model of Narrative Persuasion. Mahwah: Lawrence Erlbaum Associates; 2002.

79.   Green M, Strange J, Brock T. Narrative Impact: Social and Cognitive Foundations. Mahwah: Lawrence Erlbaum Associates; 2002.

80.   Vorderer P, Wulff H, Friedrichsen M. Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations. London: Routledge; 2013.

81.   Buckland W. Creating a cliffhanger. In: Morris N, editor. A Companion to Steven Spielberg. Hoboken: John Wiley and Sons; 2017:122-136.

82.   Bruner J, Haste H. Making Sense (Routledge Revivals): The Child's Construction of the World. London: Routledge; 2010.

83.   Wardrip-Fruin N, Mateas M, Dow S, Sali S. Agency reconsidered. 2009. URL: http://www.digra.org/wp-content/uploads/digital-library/09287.41281.pdf [accessed 2019-05-01]

84.   Bassiouni DH, Hackley C. 'Generation Z' children's adaptation to digital consumer culture: A critical literature review. J. Cust. Behav 2014 Aug 25;13(2):113-133. [doi: 10.1362/147539214X14024779483591]

85.   Wolfe CR, Dandignac M, Reyna VF. A theoretically motivated method for automatically evaluating texts for gist inferences. Behav Res Methods 2019 Jul 24. [doi: 10.3758/s13428-019-01284-4] [Medline: 31342470]

Abbreviations

AVG: active video game
MVPA: moderate to vigorous physical activity
PA: physical activity

XSL•FO
RenderX

Original Paper

Treating Children With Speech Sound Disorders: Development of a Tangible Artefact Prototype

Joaquim Santos[1], MSc; Mário Vairinhos[1], PhD; Luis M T Jesus[1], PhD

Escola Superior de Saúde (ESSUA), Universidade de Aveiro, Campus Universitário de Santiago, Aveiro, Portugal

Corresponding Author:
Luis M T Jesus, PhD
Escola Superior de Saúde (ESSUA)
Universidade de Aveiro
Campus Universitário de Santiago
Agras do Crasto
3810-193
Aveiro
Portugal
Phone: 351 234370284
Email: lmtj@ua.pt

Abstract

Background: A prototype of a tangible user interface (TUI) for a fishing game, which is intended to be used by children with speech sound disorders (SSD), speech and language therapists (SLTs), and kindergarten teachers and assistants (KTAs) and parents alike, has been developed and tested.

Objective: The aim of this study was to answer the following question: How can TUIs be used as a tool to help in interventions for children with SSD?

Methods: To obtain feedback and to ensure that the prototype was being developed according to the needs of the identified target users, an exploratory test was prepared and carried out. During this test using an ethnographic approach, an observation grid, a semistructured questionnaire, and interviews were used to gather data. A total of 4 different types of stakeholders (sample size of 10) tested the prototype: 2 SLTs, 2 KTAs, and 6 children.

Results: The analysis of quantitative and qualitative data revealed that the prototype addresses the existing needs of SLTs and KTAs, and it revealed that 5 out of 6 (83%) children enjoyed the activity. Results also revealed a high replay value, with all children saying they would play more.

Conclusions: Serious games and tangible interaction for learning and problem solving serve both teachers and children, as children enjoy playing, and, through a playful approach, learning is facilitated. A clear pattern was observed: Children enjoyed playing, and numerous valid indicators showed the transposition of the traditional game into the TUI artefact was successful. The game is varied and rich enough to be attractive and fun. There is a clear need and interest in similar objects from SLTs and educators. However, the process should be even more iterative, with a multidisciplinary team, and all end users should be able to participate as co-designers.

Introduction

The prevalence of speech sound disorders (SSD) in the United Kingdom is estimated to range from 2% to 25% in children aged 5 to 7 years [1]. In Portugal, where the fieldwork reported in this paper took place, it is estimated that thousands of preschool aged children (8%-11% of the total population) need speech and language therapy [2]. However, owing to budget cuts, schools have fewer professionals to intervene; therefore, the role of parents and kindergarten teachers and assistants (KTAs) is particularly important [3], as the significant effects that SSD can have later in life are well documented [4-6].

Speech and language therapists use physical media (games and assorted toys) to stimulate speech and help children to overcome

XSL·FO
RenderX

SSD [7]. A physical material has intrinsic qualities, such as weight, form, smell, or texture, and these can be used to stimulate speech production, thus supporting SLT interventions. A digital app, especially if tactile or mobile based, has the appeal of being a well-known device by many children, and this produces a sense of engagement through the use of sound, animation, and color. A multimodal approach, although in need of further research, appears to be effective in several fields of speech therapy [8,9], as well as other areas [10]. What is described in this paper is the creation of a hybrid artifact, capable of combining physical and digital media—multimodal, being used by speech and language therapists (SLTs), KTAs, parents and children in a one-on-one session or as group activity, with both children with SSD and typically developing children.

In this paper, the theoretical background will be discussed, as well as previous relevant projects and their contributions to the development of the current prototype; the methodology and the fishing game tangible interface (FGTI) prototype, its functional design, and technical requirements; the prototype development phase (ie, the parts that make the prototype and its iterations); detailed results of the exploratory test. Conclusions and future work are presented in the final section. The aim of this paper is to determine how can tangible user interfaces (TUIs) be used as a tool to help in interventions for children with SSD?

Speech Sound Disorders

SSD take the form of gaps in children's speech sound systems, which can cause difficulties in producing or understanding phonemes [7]. Children with SSDs also exhibit speech patterns and structures that should not be present in typically developing children of their age [7]. A child might use, on a regular basis, what is designated by SLTs as a phonological process, for example, *final consonant deletion* (ie, the child omits a consonant in the final position of the syllable or final position in the word) [11]: The Portuguese word <*porco*> (in English, <pig>) is produced as <*poco*>.

Role of Parents and Kindergarten Teachers and Assistants' Roles

The current recommended speech and language therapy practices point to a family-centered intervention, promoting not only the parents' involvement in the sessions and in homework activities but also in planning a session and setting goals. Family-centered guidelines integrate the whole family as a client, positive family and professional relationships, parental decision making, and the empowerment and enablement of families [12].

KTAs are of great importance to child development because of the time they spend with children and the nature of their relationship. They are part of a child's innermost circle [10] and can help in the detection and reporting of possible cases of SSD, as well as in the implementation of specific activities with a child, as long as proper training, support, and tools are provided to them by SLTs. KTAs are well aware of the cognitive and social impacts of SSD in children and the negative attitudes people tend to have toward them [13]. However, a caregiver must attend the needs of several children, and in Portugal, activities have to be group based and have to benefit all.

Tangible Artefacts

Beyond conventional interaction paradigms, such as graphical user interfaces or command line interfaces, there are several interaction paradigms (eg, natural interaction, ubiquitous computing, pervasive computing, mixed realities, or wearable computing) that can incorporate the activity context in an effortless interaction approach. The role that tangible user interfaces (TUIs) can play in education and health, the concept of interaction and how it differs from adults to children, and some psychological aspects that affect how children learn are briefly discussed in this paper.

Tangible User Interfaces

TUIs seek to move away from the generic combination of screen, mouse, and keyboard interaction and attempt to transform the world itself into an interface [14]. They can be defined as interfaces that support users' direct interaction with the digital world or digital device by use of real-world objects or tools [15]. They use physical forms designed and improved over millennia to fit a specific task [16], facilitating the user's discernibility and direct manipulation of the interface through the user's peripheral senses (eg, touch or vision), because of its physical embodiment [14,16]. The user can focus his or her attention and consciousness on the task and not on the interface [17]. According to Norman [18], people develop throughout their lives a process of uninterrupted adaptation to and with the environment and an understanding of how to act in the physical world. It is from this seemingly innate understanding from which the concept of *affordances* stems. Affordances, according to the original definition by Gibson [19]—particularly the affordances of the environment—are what the environment offers the animal, whether for its welfare or unease. The affordances theory lies at the center of the conceptual model of TUIs, as the incorporation of digital technology into objects of the physical world will make the interface more familiar and easier to understand from the user's viewpoint. TUIs can be approached as rigid discrete interfaces that use certain objects or shapes with which the user would interact and which have a perceived meaning with a finite set of objects and possible interactions [16], or they can be perceived as a more "organic" and material malleable, taking advantage of new digital and physical materials that can seamlessly pair sensing and display capabilities. These interfaces have the potential to break the boundaries of predetermined interactions [16].

Tangible Artefacts in Education and Health

Tangible artefacts have long been used in interactive games in therapeutic contexts, especially in fields of cerebral palsy or poststroke recovery [20,21]. They also allow one to assess several physiological parameters, without any stress associated with a visit to a doctor's office, relieving an anxiety felt by many children and some adults alike [21].

In education, both the needs of the teachers and the needs and curricula of the students have to be fully understood and satisfied [22]. Serious games and tangible interaction for learning and problem solving serve both teachers and children. Children enjoy playing, and, through a playful approach, learning is facilitated [23]. Tangible artefacts by nature invite collaboration,

allowing several users to interact with the artefact and themselves [23], thus increasing productivity levels [21], particularly as TUIs provide an interface that is space multiplexed instead of the time-multiplexed interaction, as we can typically find on conventional digital interfaces that rely on mouse and keyboard. The way in which the user interacts has to be driven through affordances, mappings, and game logic to ensure reliability and take full advantage of the potential of the artifact [21,24].

Designing Interaction for and With Children

Designing for interaction is all about how to design for people, their needs, emotions, and intellect, making it imperative to be highly aware of what to expect from those who will interact with the final product [25]. With the shift toward participatory and ethnographic methods, those designing interactive apps or objects have to fully understand how and why people use technological innovation [22,25].

The children (aged 3-7 years) targeted with the artefact created during this project are preliterate. With short attention spans, they have difficulty conceiving abstractions, and their fine motor skills are not yet fully developed [22,26]. Nonetheless, designing in a way that can be perceived as too childish can be felt to be boring or disrespectful by the children [22], as they are acutely aware of their capabilities [27]. A workaround is to embrace designing with children as co-designers, evaluators or subjects, or a combination of these. This approach has its own drawbacks, requiring that adults and children must work together, but in the end, this method assures that the design meets the needs and specificities of children [22].

How Do We Learn?

Children search for multimodal stimulation, which consequently encourages their physical and cognitive development. They are naturally motivated to explore what is around them by engaging with their environment, their medium and substance, which consequently affords certain immediate perceived venues of action, manufacture, and manipulation [19], reinforcing learning through the dynamics of play.

Cultural and social contexts influence how children construct the world and their knowledge [28]. Social experience is a critical factor in mental development [29]—interpersonal connections and social interaction provide the means for a child to access experiences that they can then integrate into their view of how things work [28]. Play helps a child to separate the meaning of an object from the actual object, and from a child's point of view, it is not just a game, it is a serious thing, which they consider as work [28]. Adults, who make up a large part of our stakeholders, also benefit from something playful and fun. As Donald Norman says in Emotional Design: Why We Love or Hate Everyday Things [30] (page 103), "Beauty, fun, and pleasure all work together to produce enjoyment, a state of positive affect." These positive emotions, as the author says, have many benefits and are pivotal in our ability to learn.

Gamification

The prototype presented in this paper is a conversion from a game, which, at first glance, does not need extra elements to be perceived as such. However, some aspects of it can be further improved, for example, a leaderboard that, when visible to all players, stimulates healthy competition. Gamification can be defined as making use of elements typically found in a game in a nongaming context to transform every day, uninteresting tasks, into engaging ones, while increasing user activity and retention [31-33].

Typical game elements with extensive use of gamification and with interest in the physical and digital part of the FGTI prototype include the use of points (and point systems) and the existence of levels. According to Zichermann and Cunningham [31], points are a vital element, and they should always be present at any gamified activity, if not in a visible way, at least visible to the activity designer only, so he or she can assess how the users interact with the activity and design appropriate outcomes [34]. When visible, points allow the user to know how close he or she is to his or her goal, and points can thus be highly motivational [34]. Levels, as the name implies, mark something, in this case they mark in-game progress, and they allow players to be aware of where they are, over time, in relation to the game experience. Levels should be logical in terms of level progression, and they should be easy to add to [34]. By further hiding away the test or activity behind a game-like approach, the stakeholders might feel more relaxed and willing to participate [32].

Sum of All Parts

An effective TUI, usable in SSD intervention, should allow some degree of simulation and storytelling, as well as the construction of mental models of knowledge. It should also provide some form of social interaction with the artifact and the other players, all in a playful atmosphere [29]. A TUI can provide natural interaction without emphasizing any cognitive effort—a child does not need to learn or understand a set of rules or settings. The perceived focus is on the action executed and what it can represent. A TUI can help gamify speech and language therapy intervention with a child; it can provide an alternative means to promote children's speech production; it can help parents by being a "fun" homework exercise to do together and can help KTAs by being an activity that can be developed in a group of normal children and children with SSDs. The prototype's intended use is at the clinic, the kindergarten, or at home. However, there are challenges to overcome. The prototype must offer more than a traditional game. It has to be lightweight and easy to transport while remaining durable to withstand daily use by children. The software component should incorporate options to intervene in several SSD and remain interesting to play with, while sending data to a log that SLTs can consult later.

Related Work

Some examples of good practices or cases of success can be found in the literature [35-39], but not one is an exact fit in terms of technological requirements, target population, or intervention area of the current project. A total of 3 projects were considered relevant for the conceptualization of FGTI: first, the *table-to-tablet* (*T2T*) intervention materials, designed to be a reliable and valid solution [40] to be used by Portuguese SLTs when treating children with SSD. It has a physical and a

digital version, and SLTs can use them interchangeably, but one does not communicate with the other. Second, the *LinguaBytes* materials, from the Netherlands [41], comprises a full set of exercises and varied activities that are mediated by tangible artefacts. The aim of *LinguaBytes* was to be a tangible language learning system for toddlers with some form of motor disability. Third, *Jabberstamp* [42], developed by a team at the MIT Media Lab (Tangible Media Group), is a tool that allows children to add sound to their drawings, collages, or paintings, enabling them to communicate more effectively before developing or mastering writing skills.

Relevance and Motivation of the Fishing Game Tangible Interface Prototype

It can be argued that the traditional fishing game is already engaging and in use in intervention by SLTs, similar to other traditional games, such as Bingo or wooden blocks, with the alphabet written on them or animals [43,44]. However, the TUI artefact can present additional advantages:

- It allows the customization of sprites (both the avatar representing the player and the fish). This customization stimulates user engagement and makes each game unique.
- It facilitates the process of preparing the session (SLTs, parents, or KTAs).: The system will set up almost everything.

It eases the burden of certain game-related tasks, such as keeping and updating a score or showing who is winning on a leaderboard. In addition, the software can introduce extra challenges, bonuses, and "power-ups."

It allows the introduction of extra gamification elements.

It affords extra motivation by presenting a game in a new format that is flexible and allows for uniqueness in each play, potentiating children's preferences toward digital games [42], while retaining the physical traits.

It can also ease the postintervention process. SLTs do not need to record any data from the session, as a "log" file will be created for them, with all the relevant information needed (player's names, age, intervention time, and what were the answers of all children). This file can (ideally) be accessed through a Web-based software or emailed to the SLTs.

The prototype, at its core, is the traditional fishing game that so many know. Conceptually speaking, the users can expect to find the same organization, functionalities, and set of rules. As such, previous experiences with a traditional fishing game will allow players to seamlessly use the prototype with just a very quick explanation of some components and their functionalities.

The prototype aimed to be innovative and solve a real-world issue, involving different participants, with diverse roles, as can be seen in Figure 1.

Several design iterations result from this richness in feedback, different uses, and perspectives, because of an encouraged participatory culture (consumer/player active in coproduction) [45]. This constant iteration and evaluation [46] are a trademark of design-based research (DBR). DBR is capable of producing 2 different and nonexclusive outcomes [47]: theoretical (this paper) and practical (the FGTI).

Figure 1. Types of users, their permissions and possible actions within the prototype.
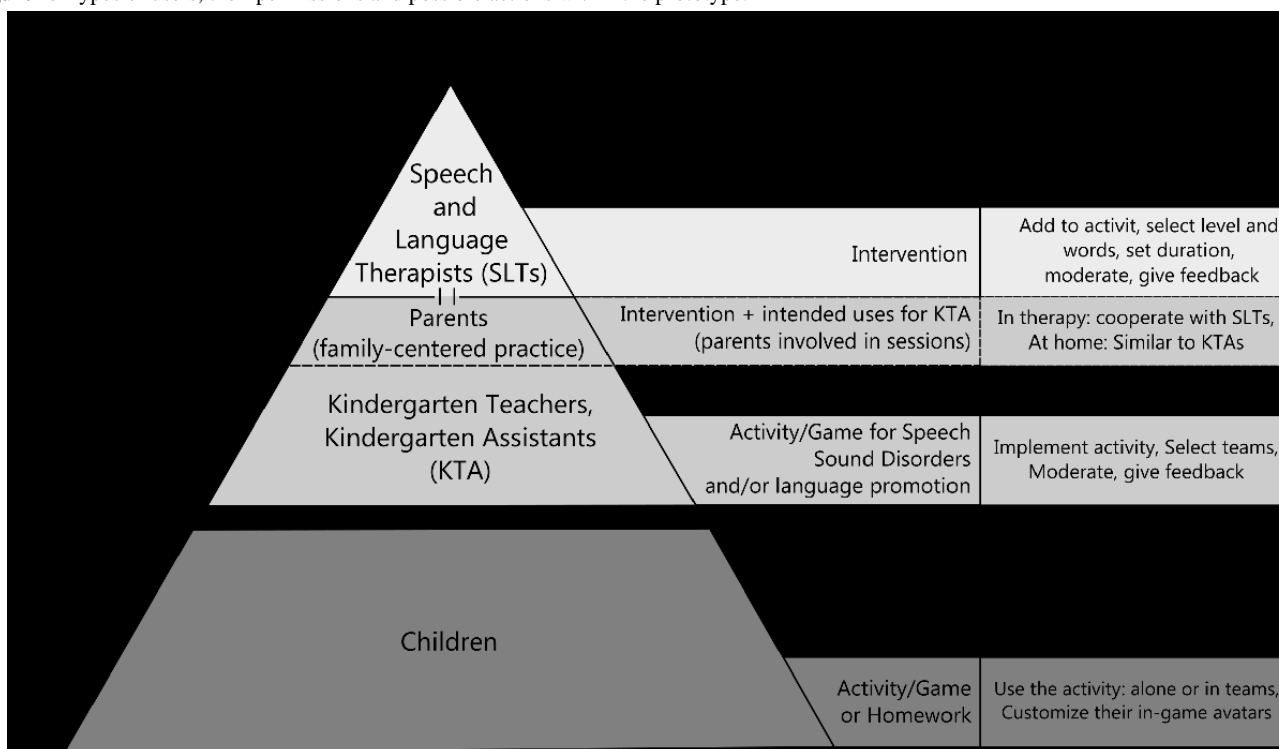
Methods

Design-Based Research

DBR allows the researcher to be involved in a way that he or she may glimpse unexpected uses or interactions with the prototype, causing a need to alter it or re-assess the target users, what they do and their needs. To reach the test phase, the physical prototype went through several revisions, always analyzed by SLTs and other project participants, and reworked accordingly. The software part of the prototype (both the game and the web app) was equally revised and improved. To gather data (qualitative and quantitative) from the designated users, in a real-life scenario, an ethnographic approach was used, and a user exploratory test was conducted regarding prototype use. These data were collected through observation and a questionnaire.

Sample Definition

The sample for the exploratory test comprised 10 expert users, with ages ranging from 4 to 55 years. The parents' informed consent was procured, in agreement with the World Medical Association's Declaration of Helsinki regarding human experimentation. Parents also received a document, briefly explaining the test. In addition, ethical permission was obtained from an independent ethics committee (Comissão de Ética da Unidade Investigação em Ciências da Saúde – Enfermagem da Escola Superior de Enfermagem de Coimbra, Coimbra, Portugal), process number P159-05/2013.

The sample can be further clustered into 3 subgroups as follows:

- Children: A total of 6 children, with ages ranging from 4 to 6 years. This group had 2 boys and 4 girls, all speakers with normal development, – no SSD.
- Speech and language therapists: A total of 2 SLTs, with ages ranging from 42 to 54 years. Both are also lecturers at the School of Health Sciences (ESSUA), University of Aveiro, Portugal.
- Kindergarten teachers and assistants: One kindergarten teacher, aged 55 years, and 1 one kindergarten teaching assistant, aged 51 years.

This sample allowed testing of a variety of situations, namely speech and language therapy intervention, children's use of the activity as a game (group activity), and KTAs with children. The only missing element(s) from the expected users were parents.

Data Gathering

The technique used was direct observation, although the instruments for data gathering were a form (qualitative data) and a semistructured questionnaire (which allowed the collection of both qualitative and quantitative data). In the creation of the form as an element to annotate the observations, great care was taken in not only dividing the observable actions but also in transforming and categorizing the observable world into interpretable and observable data.

Exploratory Test: Observation Form and Open Questions

To carry out an effective, direct, and nonintervening observation, especially of an activity involving children and a certain amount of play, a simple and easy-to-complete form was created. A set of open and closed questions, using a visual Likert like scale (a Smileyometer as shown in Figure 2) [48], was also prepared to be used at the end of the test, with the target users (children).

Owing to the variety in the sample and to the constraints an observer may face, a form was prepared to address all scenarios in a single 2-sided A4 page. It was up to the observer to know what he or she was observing and where to annotate it. Both the form and questionnaire used a unique ID for the person observed interacting with the prototype, the type of user he or she was, age, duration of the session, and date.

The form was divided into dimensions or broad areas with clearly defined parameters, to mark as observed ("yes" or "no"), as well as an area reserved to take some quick notes. Those dimensions were as follows:

- Game/Prototype Usability: Parameters regarding the ability to identify the game, its objectives, and components.
- Game/Prototype (Physical) Characteristics: Parameters revolved around the materials, colors (or lack of), robustness, and feedback from the game.
- Gamification: Parameters regarding the desire to play more, if players know when their turn to play is and their score.

The questionnaire was also set on an A4 page, and the observer would choose what to ask and to whom. It was divided into 2 parts, 1 part aimed at the children and 1 part aimed at the SLTs and KTAs. The part aimed at the children had 2 questions, with 2 Smileyometer scales—1 with 5 smiles representing values 1 to 5 for the question "Have you enjoyed playing this game?" The other used 3 smiles, representing values 1 to 3 for the question "Would you play this game again?" The remaining questions were open questions. The last part also had a set of open questions, used to conduct an informal guided interview.

Figure 2. One of the Smileyometers used.

Location and Setting Up

The test took place in a kindergarten near the University of Aveiro in Portugal. This setting allowed having, in the same place, 3 of the 5 intended users: the children, the kindergarten teachers and the kindergarten teaching assistants. The fourth intended users, the SLTs, joined the group at the kindergarten. This is also coherent with some SLTs' intervention locations—kindergartens and schools—allowing the observer to be in the intervention environment instead of a laboratory.

The test was held in the kindergarten library, as it is a quiet and spacious room with plenty of natural light. The prototype was prepared; its contents were set up, and the fish basket was connected to the laptop. A set of printed game rules was available, and the SLTs and KTAs were encouraged to read them. A brief explanation was given of how the prototype worked, the role of the components, and functions of the buttons. This explanation was considered to be similar to the instructions one would receive when buying such equipment. The observer prepared the observation forms and the set of questions to ask the users.

Four Possible Use Cases Tested

The prototype's hardware and software limitations constrained its use to 2 participants at a time. Owing to the number of participants on the exploratory test, more than one expected use scenario was tested.

The first scenario tested was the SLT intervention on a child, using the activity. The second was the kindergarten teaching assistant and a child and their interaction with each other and the activity. The third scenario was a group activity—2 children playing the activity, talking with each other about what the other caught, points won, and related subjects. The fourth and final scenario tested was a child and the kindergarten teacher playing and how both interacted. The only untested scenario was that of parent and child playing the activity because of time constraints. This limitation is discussed at the end of this paper.

Fishing Game Tangible Interface

The design and functionality of the FGTI is addressed in this subsection, as well as a description of its functionalities and needs, as related to SSD. The hardware and software used to build the prototype and game are briefly described, concluding with a synopsis of the ideas that led to the finished FGTI prototype.

Functional Design

The Fishing Game (or Pond), is a dexterity game, and the rules (which may vary from publisher to publisher or can be determined by the players) are as follows: all sea creatures and treasure chest, seen in Figure 3 in both physical and digital counterparts, go into the pond. The players take turns in attempting to catch a fish with the (magnetic) fishing pole.

Figure 3. Digital and physical activity assets.

Each player has a set number of attempts to try to catch a fish, without looking into the pond, to better emulate a real fishing activity at a lake or ocean. The number of allowable attempts is agreed upon among players at the outset. Each sea creature has a number printed on the back. For older players, the value of the sea creatures is added up to determine the winner. For younger children, the number of caught sea creatures determines the winner.

The TUI prototype rules are broadly the same as described above, with the following exceptions:

- All sea creatures go into the wooden trunk that serves as both a carrying space and the "board" game area.
- Sea creatures will have a certain range of values. For example, a codfish value can go from 5 to 15 points, mimicking the fact that the fish can be a small or a bigger codfish. How much each sea creature is worth will be

calculated randomly within a range of values, during the game.

In Figure 4, an infographic with the relevant functional components of the FGTI is shown. The solid lines represent interaction, whereas the dashed line represents the visual and auditory cues the player gets from the laptop. The dashed circle and line with the fish represent what is inside the box without perspective skew.

The stakeholders interact with the fishing rod, which in turn interacts (catches) with the fish (see Figure 4). The stakeholders interact with the captured fish by placing them in the fish basket and pressing the necessary buttons to execute the activity (differentiating the word and image) and change player's turn—the sound cue is repeated at given intervals until the stakeholder acts. The fish basket interacts with the laptop by sending and receiving data according to the fish radio frequency identification (RFID) data or activity moment. The users have constant digital (via screen) or real (via the fish basket, sea

creatures, and fishing rod) feedback, expressing the action, points received, or player turn.

The activity screens are shown in Figure 5. The stakeholders are greeted by the name of the activity (top left) and have a chance to select the level or have it chosen by the SLTs (top right). Immediately below these images, there is a representation of the visual differences between levels (calm and sunny sea for level 1 and a stormy sea for level 2). The sound effects also add to this atmosphere. The remaining images exemplify the in-game screens, the different catches, and associated messages.

The FGTI prototype, with all its components, the activity software, and the Web-based app, seen below in Figure 6, were developed to be as close as possible to the traditional game and to use the rules described above. In the case of the website or app, it was developed to act as a possible replacement for the setting up of both the game or activity and the clinical file of the child.

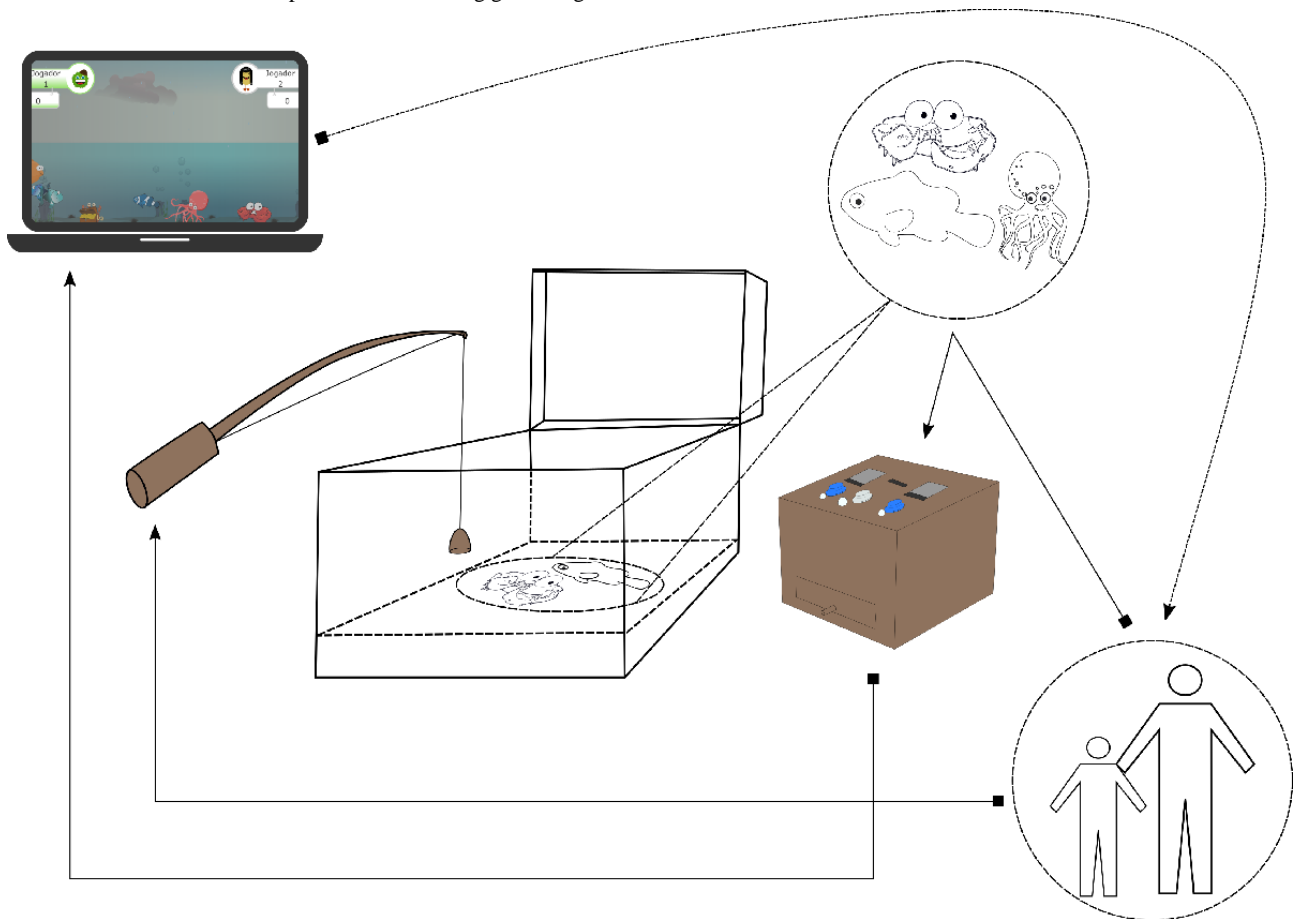Figure 4. Relevant functional components of the fishing game tangible interface.

XSL•FO
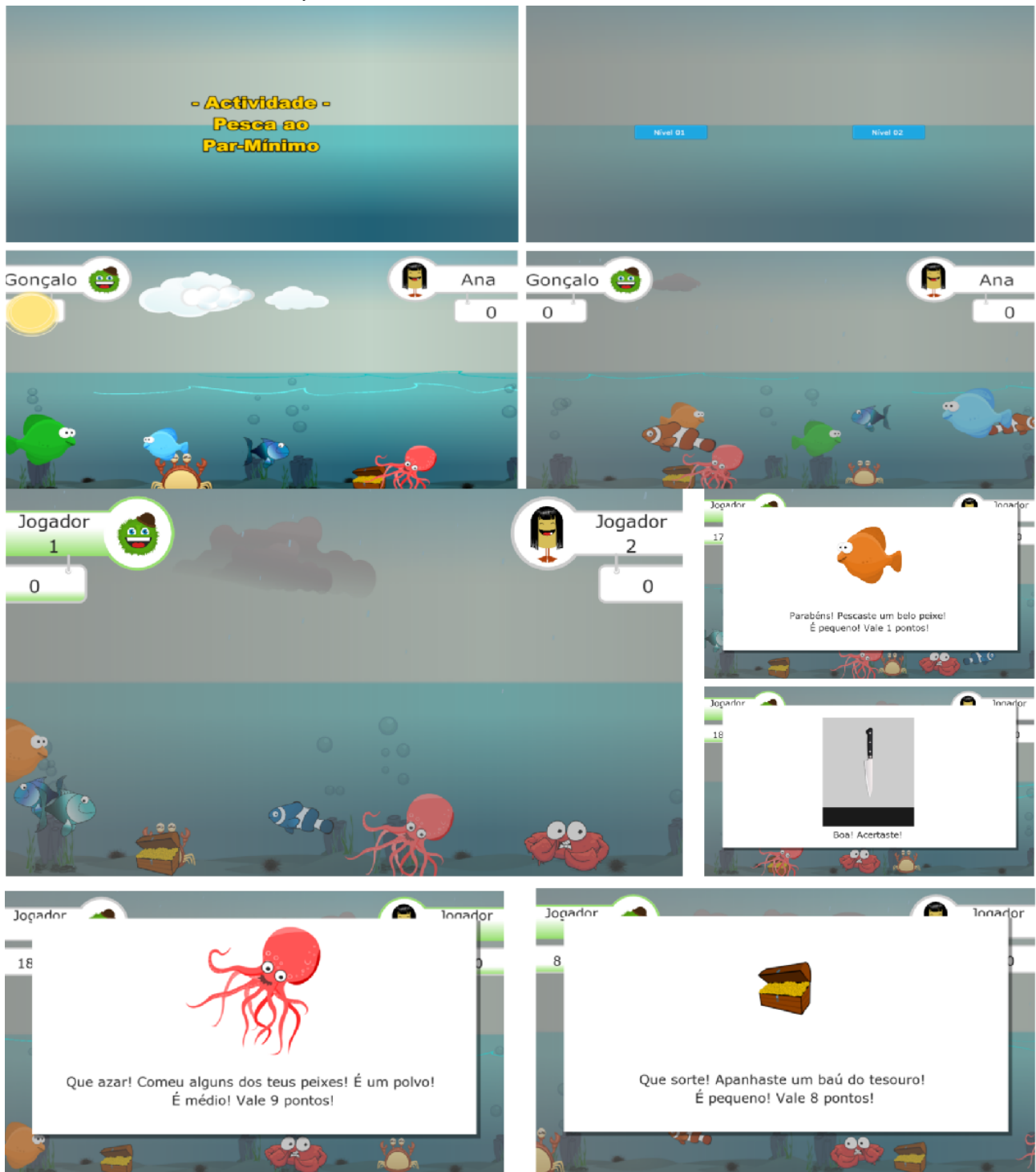RenderX

Figure 5. The different screens of the activity.

Figure 6. Web companion screen.

Technical Requirements

The FGTI was designed to be self-contained, low cost, and easy to replicate, with embedded physical computing capabilities.

Figure 7 shows the complete prototype, assembled and ready to be tested. It was planned to be easy to transport, install, and use "as is." The only requirement is the presence of an electrical outlet.

Figure 7. The prototype ready to go.

Hardware System Architecture

Owing to time and budget constraints, the liquid crystal display (LCD) screen was not placed in the wooden trunk lid. In addition, a Raspberry Pi approach was discarded for the same reasons, and a laptop (as the prototype "control center") and its screen (for display purposes) were used instead, as depicted in the block diagram in Figure 8. RFID tags glued on the sea creatures were used, and an RFID reader (RDM6300 125 KHz) communicated with an Arduino via a transmission pin, with a maximum effective (reading) distance of up to 50 mm, taking less than 100 ms to decode the tag or card. An external antenna was placed around the fishing basket slit to read the sea creatures' RFID tags.

Figure 8. Technological and software requirements diagram.
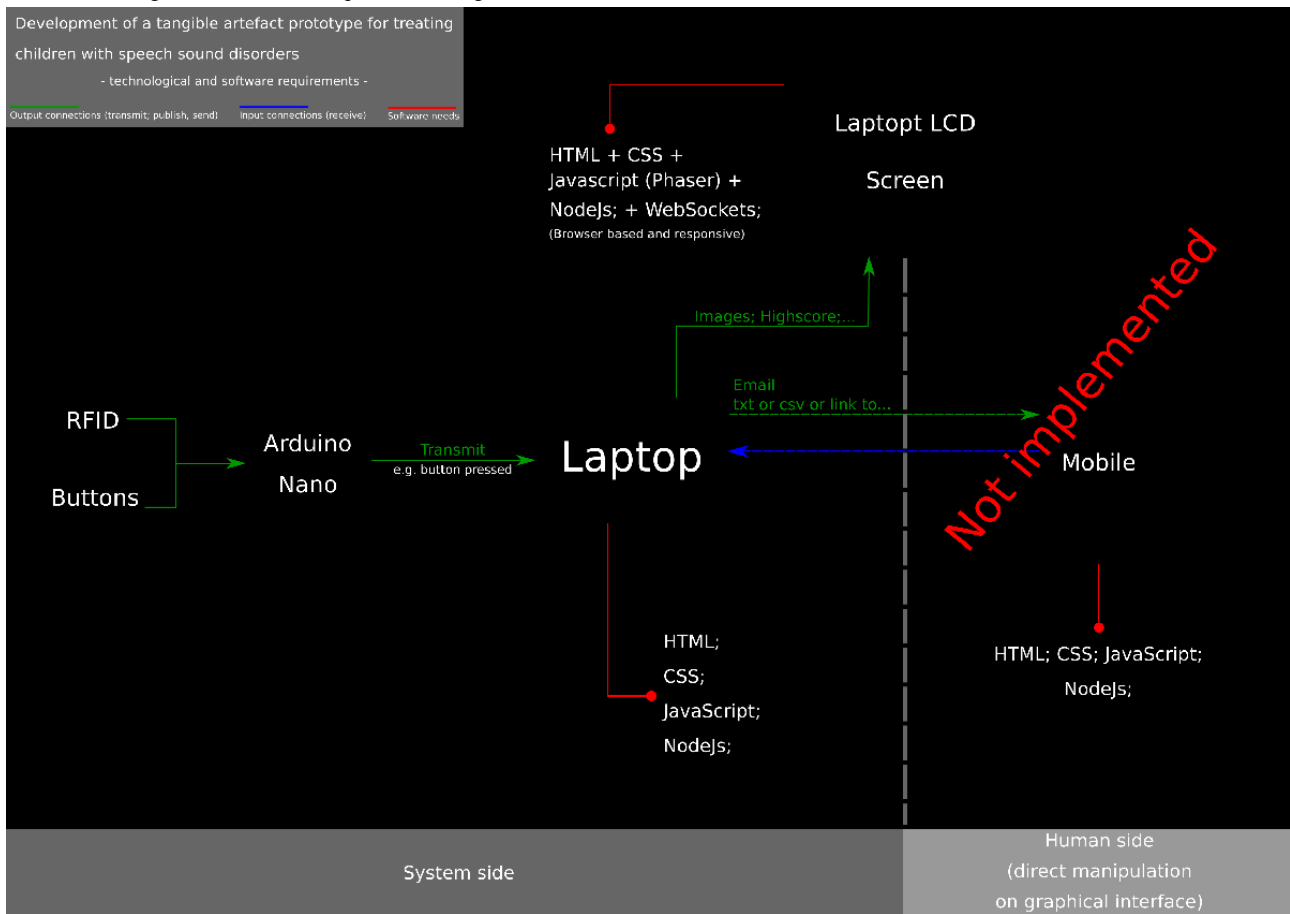
Software

The development of this prototype (activity and website or app) involved several programming languages and libraries, and various pieces of software (or scripts) were created.

The activity was coded using a mixture of HTML, cascading style sheets (CSS), and JavaScript, with the Phaser Framework (available at phaser.io.). Phaser is a JavaScript framework for 2D game development in mobile and desktop environments, also ideal for the prototype goals [49]. JavaScript Object Notation (JSON) was used to store several values and properties attributed to each sea creature, for example, name, value (range of), or the sound it represents. To make the activity more scalable and readier to deal with real-time use and requests, Node.JS was also used. Node.JS can use and run a JavaScript server side (traditionally an area for other languages, such as Hypertext Preprocessor or PHP). With Node.JS and its node package manager, several packages were installed to allow bidirectional communication between the activity and the Arduino with the server, via WebSockets (available at socket.io) plus node-serialport (available at serialport.io), or to render the HTML code in the browser through Node (express.js; available at expressjs.com). This installation implied the creation of a simple server-side and client-side set of files to handle requests and responses. The website or app companion used almost all of the previously mentioned languages and technologies, with some exceptions: The Phaser framework was not used; Node.JS was used, but with different packages installed; HTML and CSS played a larger role; Bootstrap (available at getbootstrap.com)

was heavily used. Bootstrap is a front-end library that allows one to quickly build or prototype responsive (ie, which can adapt to any screen dimension) websites or apps [50]. The website or app companion can be transformed into a mobile app installer package, with software like Apache Cordova (available at cordova.apache.org). Cordova wraps the code into a native container that can access the mobile device's native functions (such as the accelerometer) and several different platforms, enabling Web-based software to deploy in any device (platform agnostic) [51]. The software on the Arduino was coded using the Arduino programming language. A specific library was also used to integrate the RDM6300 RFID Reader. Figure 8 presents a diagram of the technological and software requirements, as well as the parts comprising the prototype.

Prototype Development

The prototype comprised the following parts: wooden trunk, sea creatures, fishing poles, fishing basket, and speakers. Some elements of the prototype, because of their importance, role, or particular challenge, required more than one revision.

Wooden Trunk

The design of the wooden trunk prototype, similar to all the elements in the fish game, was based on a maritime theme; therefore, the idea was that it should mimic an old treasure chest. That design was reflected not only in the shape (no round top lid) but also in the choice of ironmongery and straps. It encloses all the components used in the activity, except for a laptop. This

laptop is needed to power the Arduino Nano and electronics and to display the digital representation of the fishing activity.

### Sea Creatures and Treasure Chest

In this prototype, sea creatures have an extra dimension (they are not just an object with a value assigned to it or being counted as one more fish caught by the player) because of their RFID tag. This tag gives users access to as much information as needed, included in the JSON file, read during the activity. Sea creatures and the treasure chest were revised once. Similar to the treasure chest, during the build and actual use, some notes were taken on possible improvements. A total of 9 different sea creatures (6 fish, 2 crabs, and 1 octopus) and a treasure chest

were created. They were then laser cut, and the result is shown in Figure 3.

Fishing Basket

The fishing basket, shown in Figure 9, was the prototype element that took the longest to build and develop, mainly because of the fact that it is the most complete and the one closest to the initial idea of how it should be and behave. Owing to its importance and central stage role, it was the driver of several design, technical and material changes. A consideration was always taken as a decisive factor: the end result had to be cheap and easy for any hobbyist to develop.

Figure 9. The final iteration of the fish basket, from prototype (top-left corner) to the final result.
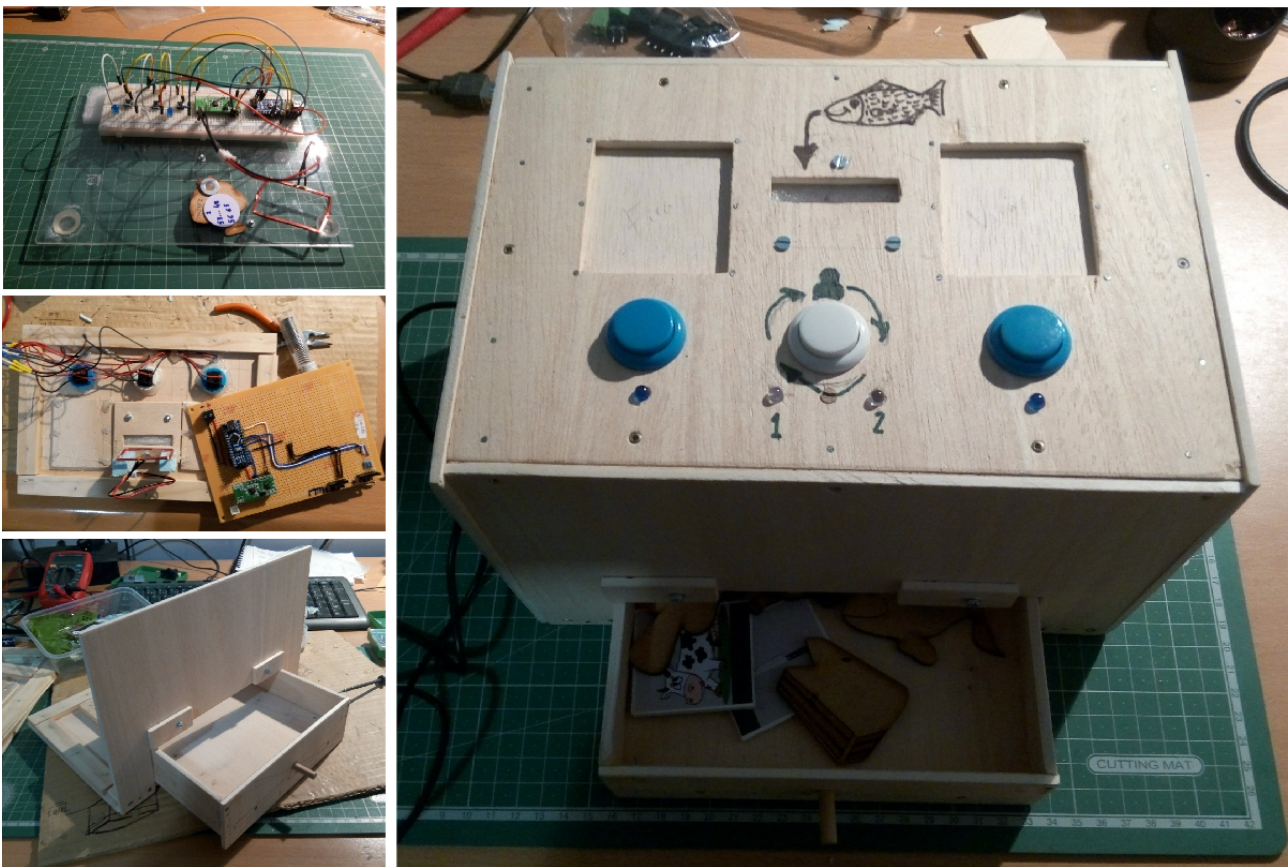
Speaker Rocks and Fishing Poles

The speakers were initially conceived as being encased in the wooden trunk lid, close to the LCD. With a reduction in size of the wooden trunk, the position of the speakers had to be reconsidered. To keep the maritime theme, the idea was to turn the speakers into a rock-like element, similar to those found around piers.

The fishing poles were revised twice (ie, they were built during the initial construction of the wooden trunk and the wooden trunk revision.02). Despite the several ideas resulting from the initial brainstorming sessions, time-related constraints led to a pragmatic and simple approach. The ideas for a more advanced fishing pole were collected to be used in a future iteration of the game.

Results

Overview

In this section, the results obtained in the exploratory test are presented and appraised. Some items were not possible to observe, or it was not possible to overhear comments about them among the users. However, they were kept in this section (see tables in the following subsections) because of 2 main reasons: first, being an exploratory test, done to ascertain the pertinence of further studies [52], the authors felt that its absence would skew the test results, and second, as some unobservable items were expected to elicit some sort of feedback (eg, the lack of color on the figures), and as such, the absence of feedback was considered relevant.

Observation Form

The observation form was divided into 3 dimensions: prototype or activity usability, prototype or activity characteristics, and gamification. It was used for direct observation, without any intervention during the activity. The observer strived to annotate all the interactions and relevant exchanges of commentaries. In the subsections below (in table format), we will present the feedback collected from every area and parameter.

Prototype Usability

This dimension explored parameters regarding the users' ability to identify the game, its objectives, and components. Results are shown in Table 1.

Table 1. Prototype usability observation parameters.

Observation parameters	Observation results
Was the player able to identify and recognize the game and its objectives?	• A total of 2 out of the 6 children tested immediately identified the game. The others were not vocal enough to demonstrate whether if they knew what they were playing, and the SLTs[a] (who remained present for the whole duration of the test) or the KTAs[b] identified the game for them. • They all knew the "classic" objectives and assumed that the novel elements (treasure chest, octopus, and crabs) had a similar value and role as the fish. Only 1 player and 1 Speech and Language Therapist noticed that the unusual elements had different score behaviors. • A player even said that she liked this game more than the classic that she had at home: "The one I have at home is a blanket and we can stand on it and fish with our hands. But this one is more fun!" – [CT, aged 6 years]. When asked why, she replied that this had a larger variety of sea creatures and some stole points (this child caught an octopus and noticed that it had halved her score). She went on to say to the Kindergarten Teacher that she should buy this game for their classroom.
Was the player able to identify the game components (sea creatures, fishing poles, and wooden trunk)?	All participants were able to correctly identify the game components and even discriminated the sea creatures (saying that one was an octopus, another a crab, etc).
Was the player able to identify the game elements and its function (buttons and slit)?	• As none of the children were able to provide a viable answer, the SLTs and the KTAs role-played with them and helped them when they first caught a fish or when they needed to change players or choose a different set of words (minimal-pairs). From that moment on, the game elements and its functions were a learned behavior. • The test that involved 2 children playing against each other was also mediated by an SLT; therefore, the same explanations/roleplays were present.
Were the game elements timings correct?	Unfortunately, it was not possible to register this parameter. The children were having such fun while using the prototype and the SLTs and KTAs were so involved with them that none seemed to notice (or care about) the timings needed for a fish to be recognized or for the word to appear. Further testing is required to understand whether the timings are correct. One cannot assume that the stakeholders will be this engaged all the time and must instead assume that the novelty of the situation and the fact that it was a "one-off" test made stakeholders unaware of the timings.
Did the player know when it was his or her turn to play?	Regarding the entire sample, 9 out of 10 knew when his or her turn to play occurred. However, this does not mean that they took the correct steps to play or pass their turn. They all knew that after player X would be their turn. But player X would usually forget to press the button to change player, or the next player would fail to realize the player number mismatch and correct the situation by pressing the button. This would result in points being given to the wrong player.

[a]SLTs: speech and language therapists.

[b]KTAs: kindergarten teachers and assistants.

Prototype (Physical) Characteristics

Parameters in this dimension are those regarding the game components, their physical characteristics, and whether they served their intended purpose. Many of these parameters (see Table 2) were observable only in the children. This limitation was because of the brief explanation given to the SLTs and KTAs, which made them aware of the uses and whys of many of the physical characteristics.

Gamification

In this dimension, the parameters observed were those regarding the desire to play more, whether the player knows if or when is his or her turn to play and his or her score. Results are presented in Table 3.

Table 2. Prototype physical characteristics observation parameters.

Observation parameters	Observation results
Were the game and its elements adequate?	No comments from the users were registered.
Was the color of the game elements adequate?	There were no comments about the (lack of) color of the physical components of the activity, and little attention was paid to the on-screen elements.
Do the materials used to build the prototype invite the handling of it?	No comments were registered. The observer marked the parameter to ask in the guided conversation with the SLTs[a] and KTAs[b].
Was the prototype considered robust?	No comments were registered. The observer marked the parameter to ask in the guided conversation with the SLTs and KTAs.
Was the feedback throughout the activity efficient?	A total of 2 out of the 6 children waited for the feedback (eg, the audio feedback of the word after inserting the sea creatures into the slit) or were aware of it.
Did the physical constraints serve their purpose?	This was observable in 3 out of the 6 children. It was more apparent regarding the slit and its use.
Was the mapping of the buttons and their actions consistent and correctly perceived? Were they used during the activity?	A total of 4 out of the 6 children correctly and consistently used the buttons when they were supposed to and to the desired end.

[a]SLTs: speech and language therapists.

[b]KTAs: kindergarten teachers and assistants.

Table 3. Prototype gamification observation parameters.

Observation parameters	Observation results
Were the participants able and willing to play or participate until the end?	All the participants were involved until the end, showing great interest and willingness. Some (children included) even wanted to know details about the study.
Were the participants willing to play more?	A total of 4 out of 6 children asked whether they could play more.
Were the participants aware of their score, at any given time?	A total of 3 children knew their score, 1 child was not aware of it, and for the other 2 children, this was unknown. Despite knowing the score and being quite attentive to the value of each captured sea creature, the children did not seem to have a clear notion of who was winning. They looked at their points on the screen and would say, "I have x points!" but nothing more. When the activity ended, children would often ask who had won.

Open Questions Questionnaire

The open questions questionnaire was divided into 2 sections: 1 section for children and another section for SLTs and KTAs. Below, we discuss the results and answers obtained in both sections.

Children

Children were asked 4 questions, 2 of which using Smileyometers [48], with 2 scales: 1 scale had 5 smiles; therefore, values would range from 1 to 5, whereas the other scale had 3 smiles, with values ranging from 1 to 3. Children being our target users, their feedback was very important; therefore, the observer was very attentive to what they said about the activity. The *first question* was whether the participant had enjoyed playing the activity. The Smileyometer scale was explained to the children in the following manner:

> *You see, the first smile is sad. He has not enjoyed playing this game. The second smile is –well,...I really do not care about it. The third smile liked it a bit. He could play more...or not! The fourth smile is happy, he enjoyed playing the game. The fifth smile is really really happy because he enjoyed it a lot.*

This was said while pointing to each smile in turn. The child was then asked to choose the smile that depicted his or her feeling toward the game he or she had just played.

A total of 5 out of 6 children (83%) picked the happiest smile, which had a value of 5, and just 1 child chose the third smile, which had a value of 3. That is, more children enjoyed playing the activity than not.

The *second question* concerned what feature the stakeholders enjoyed the most. Despite this being an open question (ie, with a broad range of possible answers), most of the children mentioned similar aspects as the most enjoyable. A total of 4 children mentioned that they enjoyed capturing the sea creatures. One child remarked that doing so was like going fishing (actual fishing) with her father. A total of 3 children enjoyed pressing the buttons, and 2 children enjoyed placing the sea creatures in the slit. Some additional comments included that they had enjoyed the fishing pole itself, the crab, the octopus, the chance to play the game, and the variety of sea creatures, as contrasted with the original game using "just" fish.

The *third question* probed what feature the stakeholders enjoyed the least. This was also an open question. A total of 3 children did not identify anything they disliked. The remaining 3 children reported different aspects: 1 child disliked catching the crab as

it stole points; another disliked pressing the buttons, explaining, "...you press the button and that's it!" The third child disliked watching the game on the PC, preferring instead to play it physically. The gameplay was not fully understood by the stakeholders. Many were not aware that catching a crab or an octopus decreases a player's points. Similarly, many were not aware that answering correctly (pressing the buttons) would double their score. To overcome this issue, a better visual representation of the elements, with both auditory and physical cues, must be implemented on the existing FGTI components or components yet to be developed.

The *fourth question* regarded whether the stakeholder would play again. This was a question made while showing a Smileyometer with 3 values. The scale was explained to the children in the following manner:

> *Ok we have more smiles. You see, the first smile is sad. He does not want to play again. The second smile is again indifferent; he really does not care if he plays or not. The third smile is so happy he cannot wait to play again.*

All children picked the happiest smile (ie, all children wanted to play again).

Speech and Language Therapists and Kindergarten Teachers

In relation to the SLTs and KTAs, the open questions were introduced in an informal talk that occurred after the testing of all the children and after them having had a chance to try out the activity themselves. This talk took the form of a fluid conversation, with moments that resembled a brainstorm session. The observer mainly listened, offering his comments occasionally.

The *first question* was whether they would change anything in the concept or prototype. Some said they would not change anything, that it was interesting to see this transformation of a classic game and that it was nice to have the auditory feedback when something was inserted into the slit. Others had more, sometimes divergent, opinions. An SLT said that given children's (and their own) difficulty remembering to press the button to change players, this function could be automated. This suggestion prompted an immediate response from the other SLT, who said that some children had mentioned that they enjoyed pressing the buttons; therefore, this change in button press should be maintained. What this action needed was more feedback or to be made more visible, she suggested.

The *second question* concerned whether they would change anything in the game or activity. The Kindergarten Teacher said that she would not change anything in the game, as it is so similar to the "original" game that no explanations are needed. However, she felt that the way the score was being handled lacked more and better feedback for the player. This feedback would, in her opinion, increase the competition and game aspect of the activity among the users.

The SLTs both enjoyed the game. They would add more sea creatures, so you can have more auditory stimuli, as at the moment, there are only 3 sea creatures per player: they suggested

6 to 8 sea creatures per player. With larger number of sea creatures, they immediately added that the agents of distraction (the crab and octopus) numbers should also increase to maintain proportion and level of challenge. A sort of "superclass" sea creature, capable of shifting the entire game for all players, was suggested. The way the auditory stimulus was working was also remarked as faulty. Children should pick the sea creature, insert it, and listen to the word (in loop if needed), and only when this action is completed, should they see the image. What was happening was that the image would come up and allow the children to have a visual (not auditory) cue. This feature will be revised in future versions of the prototype.

The *third question* enquired as to how they would expand the prototype if they could. The kindergarten teacher mentioned that she works with bilingual children; furthermore, as a second language is taught at a young age, she then suggested the prototype should be multilingual. The SLTs would like the prototype to have more uses and not be just the game of fishing. They suggested to give names to the sea creatures to make children repeat those names, allowing sound production stimulation. Another approach to extend the game's functionality would be to use more game scenarios that the adult could choose from. Some sea creatures with some word values associated with them could be tossed into the pond for children to capture. When captured and inserted into the slit, the word would be unveiled, and the children would be encouraged to use it to construct a story.

The *fourth question* was a preamble to the fifth and sixth questions, depending on the answer. The SLTs and KTAs were asked whether they would like to see more games made into TUI. If they answered "yes," they would go on to question 5—"If yes, which games?"—if not, they would go to question number 6: "If not, why?" They all answered, "yes."

The *fifth question* was which games they would like to see made into a TUI artifact. The answers were as follows: Tic-Tac-Toe, Snakes and Ladders, Bingo, and an indirect identification game similar to What am I? These can all be possible stand-alone development avenues or add-ons to the existing TUI artifact, adding extra elements and adjusting others, along with the software binding them all.

Questions from the observation form, which were asked during the conversation, as it was deemed more productive, were about the robustness of the prototype and the adequacy of materials and how appealing it was to be used. Both KTAs and SLTs alike said the prototype was robust enough if the activity was used under (adult) supervision. Left unattended, it was robust, but accidents do happen. Regarding the kind of material and its appeal, they all agreed that wood is used quite often and that children are used to seeing it in other toys and enjoy using them. They mentioned that some parts could use other materials to give extra-sensory input that may foster immersion in the activity.

Speech and Language Therapists and the App

Both SLTs were shown the partially developed hybrid app—hybrid apps are web (browser based) apps, developed using HTML, CSS, and JavaScript, which are then wrapped in

a layer that allows interaction with smartphone and tablet hardware and software, independent of code or language. This app will allow one to prepare sessions remotely and provide the visible front of a back-office area where SLTs can keep and generate logs or visual representations (ie, graphics and other visual forms to represent information) from sessions and target users. SLTs can log in to access a roster of target users or add a new target user. They can choose an area of intervention, a set of exercises, what words or facilitator sounds to use, and other functionalities. This Web-based app can even be used by KTAs or parents, which will be able to access and set different parts of the application. Their feedback was collected and saved for a future implementation.

Discussion

Although it is true that in speech and language therapy, the sound, visual cues from the SLT, and speech production trials are essential, the main purpose of this game was not to replace an SLT but rather to provide a tool, capable of being used by SLTs, parents, or kindergarten staff, with a focus on hearing discrimination instead of articulation. It was felt that if a child struggles to produce a sound, the adult using the game can offer guidance, albeit without the know-how of an SLT. This particular game focused on a single minimal-pairs exercise, which comprises differentiating similar sounding words. Those words or "sound cues" were played until the user decided to press button A or B (associated with an image that corresponded to one of the sound cues). The SLT present in the room, if any, can choose to say the word to show the articulation, if needed.

By using an ethnographic approach and a single observer, present but not participating with any form of help besides the setting up, the impact of the observer presence was kept to a minimum. However, a known risk is that the observer is not entirely unbiased and his/her presence will always have an impact. In addition, the inability to generalize from the data gathered, the sample size and the analytical transparency (the observer was also responsible for coding the data) are points that will have to be addressed in future research, by the use of complementary tools to gather data. Despite these negative points, the richness of the data surmised in location allows for a better insight on the users' reaction and relation with technology and their natural environments [53,54].

The analysis of quantitative and qualitative data revealed that the prototype addresses the existing needs of SLTs and KTAs and that 5 out of 6 (83%) children enjoyed the activity. Results also revealed a high replay value, with all children saying they would play more. In the following section, we discuss the feedback collected and summarize some ideas and areas for improvement.

Observation Form

The observation form was divided into 3 dimensions: prototype or activity usability, prototype or activity characteristics, and gamification. It was used for direct observation, without any intervention during the activity. The observer strived to annotate all the interactions and relevant exchanges of commentaries. In

the subsections below, we will present the feedback, ideas, and areas for improvement, collected from every area and parameter.

Prototype Usability

The feedback and perceived recognition, as well as the ease with which the players started playing the game, are valid indicators that the transposition of the traditional game into the FGTI artifact was successful. The added value/interest of the extra sea creatures (octopus and crab) and treasure chest should be noted. However, the fact that the score was different for each object (as well as the on-screen object scale that changed according to the points earned) was not readily seen, and only 2 out of 10 participants noticed it. Better audio and visual feedback are needed in the next iteration to ensure this is an understood behavior, which can transform the gameplay experience. The game components seem to be well designed and help in identifying the game and the activity (fishing) on which the game is based. Some sea creatures can be further developed, as the stakeholders engaged in conversations about fish variety (eg, is it a whitefish or a cod?). Further testing with all the stakeholders is required to understand whether the affordances and design work and whether, within minutes, the stakeholders are ready to use the activity to its fullest. Furthermore, testing is required to understand whether the timings are correct. One cannot assume that the stakeholders will always be highly engaged in each play-through of the game, assuming that the novelty of the situation and the fact that it was a "one-off" test made stakeholders unaware of the timings. Knowing one's place in a queue, which in this case was his or her turn to play, does not seem dependent (especially in an activity with just 2 players) on any sort of visual or auditory feedback. It is something that the players do and know. However, interacting with a button to signal this change is not an expected behavior, and as such, despite knowing (SLTs and KTAs) or being told (children) and despite the markings above the button, this was generally overlooked and was a source of distress.

Prototype Physical Characteristics

Further testing is required to understand whether the dimensions are correct and to determine the importance of color. As the digital part was running on the laptop screen, off-angle relative to the wooden trunk and the main activity area, it may have been perceived as a "secondary" thing to look at. Tighter integration of the display (ie, with the LCD on the wooden trunk lid) may lead to a better understanding and ability to test this parameter. Additional testing and modifications are needed to improve on feedback, as per user suggestions and observation.

On at least 2 occasions, sea creatures were pushed into the slot so fast that the RFID reader was unable to detect them; therefore, this issue also has to be considered.

Several suggestions were given by the SLTs and KTAs to slightly alter the phrases and sounds to give more feedback regarding, for example, the score. A possible reason for half the children being aware of the physical constraints, such as the slit and its use, may relate to their turn in playing the activity. The first one to play used the constraint (and the SLTs/KTAs help) to know what to do with the sea creatures and the slit, whereas

the second did it because she or he had watched the first one do it.

The mapping of the buttons and their actions were generally well understood and used. There was some "natural resistance" against using the button to change players; the users seemed to expect it to be automatic. The 2 children who did not seem to be capable of using the buttons did play the activity and enjoyed it. The other player or the SLTs/KTAs helped them with the buttons, and they started using them after a short while.

Gamification

All stakeholders enjoyed taking part in the activity. Although some of this (especially regarding children) can be dismissed or looked at, considering the novelty factor, it is encouraging to get this feedback.

More than half stated their desire to play more. The remainder may not have expressed this desire for a number of reasons: for example, the test being run just before the Christmas holidays and children being busy taking part in dance or music rehearsals for the school's Christmas show; therefore, they wanted to go back to those activities.

The stakeholders were aware of their score, or they at least knew where to look at, on screen. However, most of them were unable to say who had won and by how many points.

Open Questions Questionnaire

The open questions questionnaire was divided into 2 sections: children and SLTs and KTAs. Below, we discuss the answers obtained in both sections.

Children

The majority of children enjoyed playing the activity. Even the child who answered differently did not choose a negative smile or value; therefore, even she would play.

A clear pattern could be observed: Children enjoyed playing the game. The game is varied and rich enough to be attractive and fun; however, the digital part of it went unnoticed. This oversight may arise from the previously mentioned off-angle screen. Further testing with a new FGTI revision (of all its components) is required.

All children wanted to play again. Even considering a factor like "novelty," the results are encouraging and very positive. What remains unclear and depends on further and more exhaustive testing is whether this willingness to play would be sustained after a number of games played.

Speech and Language Therapist and Kindergarten Teachers

A need and desire for tools, such as the FGTI prototype, were clearly observed. The prototype was seen as almost ready for serious field and clinical testing after some small bugs and feedback issues were solved. Both SLTs and KTAs said it was something that they could see themselves using with their children.

Suggestions and Improvements on the Basis of Feedback Gathered During the Exploratory Test

During the entirety of the exploratory test, the observer was able to collect much input and feedback, some direct, as in the final conversation with SLTs and KTAs, as well as some indirect, overheard during the activity testing. Those suggesting improvements and feedback are addressed in this section. Keeping with the maritime theme, a pirate object/figure could be created. This object would be able to steal all points from all players (in practice, resetting the entire score and doing "tabula rasa" of all progress) and sail away (meaning the pirate would be thrown again into the wooden trunk). This new element would be used to counterbalance the treasure chest element and would increase the challenge factor. In addition, the score could be displayed inside a starfish and could have some sort of animation to increase the visual feedback. A total of 2 sets of scores could be used. One, to reward the better fisherman, would be based on a number of sea creatures caught and how much they are worth. The other score would be to reward the player who answered correctly more times (in practice who did better at word discrimination, in the case of the minimal pairs). This reward would make being attentive and answering correctly worth more to a player. The number of points for each catch should be more evident with the use of auditory cues. In the game's present iteration, a voice can be heard saying, "You caught a fish. It's small." The voice should say, "You caught a fish. It's small. It's worth *X* points!" Similarly, the voice should also say, after some interaction from the player, (eg, a successful catch but unable to discriminate the word, catching a crab) the number of points the player has. Furthermore, what the voice says and what is written on screen have slight differences that should be addressed. For example, the voice says (in Portuguese), "Tenta de novo!" *(Try again)* and what appears written is, "Tenta outra vez!" One SLT felt that it would be interesting and would help with the immersion to see a representation of the fishing pole on screen, to have an idea of its (relative) position.

Conclusions and Future Work

The feedback, the perceived recognition, and the ease with which children started playing the game are valid indicators that the transposition of the traditional game into the TUI artifact was successful. The added interest in the extra sea creatures and treasure chest should be noted. However, the fact that the score differed for each object (as well as the on-screen object scale that changed according to the points earned) was not readily seen, and only 2 of 10 participants noticed it. Better audio and visual feedback is needed in the next iteration to ensure that this is an understood behavior, which can transform the gameplay experience. Several suggestions were given by the SLTs and KTAs to slightly alter the phrases and sounds, to give more feedback regarding, for example, the score. Regarding the SLTs and KTAs, a need and desire for tools, such as the FGTI, were clearly observed. The FGTI and what can be produced with the concept behind it can be extremely modular and versatile. However, certain aspects should be improved to have an even better product. The TUI game of "Pond," with the interest caused and ease of use, appears to be a suitable tool for the classroom or as an SLT set of intervention tools; however,

XSL·FO

RenderX

there should be more tests with users, which would produce more iterations of all aspects of the prototype (physical and software). Owing to time constraints, an important group of stakeholders, the parents, was not tested, but they must be involved, observed, and questioned in future work. They share, with the child, a very important space (home), and they spend special time and form unique bonds; therefore, they possess unheard information, not available to any of the other stakeholders. A multidisciplinary team that included children,

SLTs, KTAs, and parents as co-designers would allow the development of an end product that is suitable to respond to each user's needs and desires. More diverse means of gathering data and analyzing data should be employed to minimize the known flaws of ethnography. The FGTI's perceived value and the positive impact it may have as a tool for intervention for children with SSD should be further expanded. The monitoring part of the game, capable of producing usable reports for the SLTs, will be implemented in future revisions.

Conflicts of Interest

None declared.

References

1. Law J, Boyle J, Harris F, Harkness A, Nye C. Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. Int J Lang Commun Disord 2000;35(2):165-188. [doi: 10.1080/136828200247133] [Medline: 10912250]

2. Castro A, Caetano D. History of language impairment/developmental language disorder in Portugal. In: Law J, McKean C, Murphy CA, Thordardottir E, editors. Managing Children with Developmental Language Disorder: Theory and Practice Across Europe and Beyond. Abingdon: Routledge; 2019:374-386.

3. Jesus L, Martinez J, Valente A, Costa M. Speech and language therapy service delivery: overcoming limited provision for children. Public Health 2017 Oct;151:39-50. [doi: 10.1016/j.puhe.2017.06.013] [Medline: 28710926]

4. Lancaster G. Developing Speech and Language Skills: A Resource Book for Teachers, Teaching Assistants, and Speech and Language Therapists. London: Routledge; 2008.

5. McCormack J, McLeod S, McAllister L, Harrison LJ. A systematic review of the association between childhood speech impairment and participation across the lifespan. Int J Speech Lang Pathol 2009;11(2):155-170. [doi: 10.1080/17549500802676859]

6. ASHA. Budget Cuts: Maintaining Speech-Language Pathology and Audiology Services in Schools URL: https://www.asha.org/SLP/schools/Budget-Cuts-Schools/ [accessed 2018-11-28] [WebCite Cache ID 74GS2SgSZ]

7. Bowen C. Children's Speech Sound Disorders. Second Edition. Chichester: John Wiley & Sons; 2015.

8. Williams JH, Massaro DW, Peel NJ, Bosseler A, Suddendorf T. Visual-auditory integration during speech imitation in autism. Res Dev Disabil 2004;25(6):559-575. [doi: 10.1016/j.ridd.2004.01.008] [Medline: 15541632]

9. Rose M, Sussmilch G. The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. Aphasiology 2008 Jul;22(7-8):691-706. [doi: 10.1080/02687030701800800]

10. Sathappan AV, Luber BM, Lisanby SH. The Dynamic Duo: Combining noninvasive brain stimulation with cognitive interventions. Prog Neuropsychopharmacol Biol Psychiatry 2019 Mar 8;89:347-360. [doi: 10.1016/j.pnpbp.2018.10.006] [Medline: 30312634]

11. Jesus L, Lousada M, Domingues D, Hall A, Tomé D. Phonological processes in Portuguese children with speech sound disorders. Poznan Stud Contemp Linguist 2015;51(1):75-88. [doi: 10.1515/psicl-2015-0003]

12. Pappas NW, McAllister L, McLeod S. Parental beliefs and experiences regarding involvement in intervention for their child with speech sound disorder. Child Lang Teach Ther 2016;32(2):223-239. [doi: 10.1177/0265659015615925]

13. McLeod S, Daniel G, Barr J. 'When he's around his brothers … he's not so quiet': the private and public worlds of school-aged children with speech sound disorder. J Commun Disord 2013;46(1):70-83. [doi: 10.1016/j.jcomdis.2012.08.006] [Medline: 22995337]

14. Ishii H, Ullmer B. Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms. In: Proceedings of the ACM SIGCHI Conference on Human factors in computing systems. 1997 Presented at: CHI'97; March 22-27, 1997; Atlanta, USA p. 234-241. [doi: 10.1145/258549.258715]

15. Azuma R, Baillot Y, Behringer R, Feiner S, Julier S, MacIntyre B. Recent advances in augmented reality. IEEE Comput Grap Appl 2001;21(6):34-47. [doi: 10.1109/38.963459]

16. Ishii H. Tangible Bits: Beyond Pixels. In: Proceedings of the 2nd international conference on Tangible and embedded interaction. 2008 Presented at: TEI'08; February 18-20, 2008; Bonn, Germany p. xv-xxv. [doi: 10.1145/1347390.1347392]

17. Ishii H. Bottles: a transparent interface as a tribute to Mark Weiser. IEICE Trans Inf Syst 2004;E87-D(6):1299-1311 [FREE Full text]

18. Norman D. The Design of Everyday Things. New York: Basic Books; 2016.

19. Gibson JJ. The Ecological Approach to Visual Perception. New York: Psychology Press; 1986.

20. Li Y, Fontijn W, Markopoulos P. Tangible tabletop game supporting therapy of children with cerebral palsy. In: Markopoulos P, Ruyter B, IJsselsteijn W, Rowland D, editors. Fun and Games. Berlin: Springer; 2008:182-193.

21. Vonach E, Ternek M, Gerstweiler G, Kaufmann H. Design of a Health Monitoring Toy for Children. In: Proceedings of the The 15th International Conference on Interaction Design and Children. 2016 Presented at: IDC'16; June 21-24, 2016; Manchester, UK p. 58-67. [doi: 10.1145/2930674.2930694]

22. Bruckman A, Bandlow A, Forte A. Hci for kids. In: Sears A, Jacko J, editors. The Human-Computer Interaction Handbook. Boca Raton: CRC; 2007:793-809.

23. Almukadi W, Boy A. Enhancing collaboration facilitating children's learning using TUIs: a human-centered design approach. In: Zaphiris P, Ioannou A, editors. Learning Collaboration Technologies. Berlin: Springer; 2016:105-114.

24. Antle A. The CTI Framework: Informing the Design of Tangible Systems for Children. In: Proceedings of the 1st international conference on Tangible and embedded interaction. 2007 Presented at: TEI'07; February 15-17, 2007; Baton Rouge, USA p. 195-202. [doi: 10.1145/1226969.1227010]

25. Verplank B. Bill Verplank. 2009. Interaction Design Sketchbook. URL: http://www.billverplank.com/IxDSketchBook.pdf [accessed 2019-11-05]

26. Thomas JR. Acquisition of motor skills: information processing differences between children and adults. Res Q Exerc Sport 1980 Mar;51(1):158-173. [doi: 10.1080/02701367.1980.10609281] [Medline: 7394282]

27. Sherwin K, Nielsen J. Nielsen Norman Group: UX Training, Consulting, & Research. Children's UX: Usability Issues in Designing for Young People. URL: https://www.nngroup.com/articles/childrens-websites-usability-issues/ [accessed 2018-11-27] [WebCite Cache ID 74FinIOkY]

28. Madej K. Physical Play and Children's Digital Games. Berlin: Springer; 2016.

29. Almukadi W, Stephane L. Interactive system supporting children pleasurable learning. In: Chung W, Shin C, editors. Advances in Affective and Pleasurable Design. Berlin: Springer; 2017:67-78.

30. Norman DA. Emotional Design: Why We Love (or Hate) Everyday Things. New York: Basic Books; 2005.

31. de Vette F, Tabak M, Dekker-van Weering M, Vollenbroek-Hutten M. Engaging elderly people in telemedicine through gamification. JMIR Serious Games 2015 Dec 18;3(2):e9 [FREE Full text] [doi: 10.2196/games.4561] [Medline: 26685287]

32. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]

33. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining 'Gamification'. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek'11; September 28-30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]

34. Zichermann G, Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. First Edition. Sebastopol: O'Reilly; 2011.

35. Bakker S, Vorstenbosch D, Hoven E, Hollemans G, Bergman T. Weathergods: Tangible Interaction in a Digital Tabletop Game. In: Proceedings of the 1st International Conference on Tangible and Embedded Interaction 2007. 2007 Presented at: TEI'07; February 15-17, 2007; Baton Rouge, USA. [doi: 10.1145/1226969.1227000]

36. Girouard A, Solovey E, Hirshfield L, Ecott S, Shaer O, Jacob R. Smart Blocks: A Tangible Mathematical Manipulative. In: Proceedings of the 1st international conference on Tangible and embedded interaction. 2007 Presented at: TEI'07; February 15-17, 2007; Baton Rouge, USA p. 183-186. [doi: 10.1145/1226969.1227007]

37. Kaltenbrunner M, Bencina R. reacTIVision: A Computer-Vision Framework for Table-based Tangible Interaction. In: Proceedings of the 1st international conference on Tangible and embedded interaction. 2007 Presented at: TEI'07; February 15-17, 2007; Baton Rouge, USA p. 69-74. [doi: 10.1145/1226969.1226983]

38. Petrelli D, Dulake N, Marshall M, Willox M, Caparrelli F, Goldberg R. Prototyping Tangibles: Exploring Form and Interaction. In: Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction. 2014 Presented at: TEI'14; February 16-19, 2014; Munich, Germany p. 41-48. [doi: 10.1145/2540930.2540966]

39. Ruhmann L, Otero N, Oakley I. A Tangible Tool for Visual Impaired Users to Learn Geometry. In: Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction. 2016 Presented at: TEI'16; February 14-17, 2016; Eindhoven, Netherlands p. 577-583. [doi: 10.1145/2839462.2856536]

40. Jesus L, Martinez J, Santos J, Hall A, Joffe V. Comparing traditional and tablet-based intervention for children with speech sound disorders: a randomised controlled trial. J Speech Lang Hear Res 2019;62(11):4045-4061. [doi: 10.1044/2019_JSLHR-S-18-0301] [Medline: 31644381]

41. Hengeveld B. Designing LinguaBytes: A Tangible Language Learning System for Non- or Hardly Speaking Toddlers. Eindhoven, The Netherlands: Technische Universiteit Eindhoven; 2011.

42. Raffle H, Vaucelle C, Wang R, Ishii H. Jabberstamp: Embedding sound and voice in traditional drawings. In: Proceedings of the 6th International Conference on Interaction Design and Children. 2007 Presented at: IDC'07; June 6-8, 2007; Aalborg, Denmark. [doi: 10.1145/1297277.1297306]

43. Bowen C. Caroline Bowen Speech-Language Therapy. Controversial Practices in Children's Speech Sound Disorders - Oral Motor Exercises, Dietary Supplements, Auditory Integration Training. URL: https://www.speech-language-therapy.com/index.php?option=com_content&view=article&id=28:controversy&catid=11&Itemid=101 [accessed 2018-11-28] [WebCite Cache ID 74GP2xnuQ]

XSL•FO

RenderX

44. Gillon G, McNeill B. University of Canterbury. 2007. Integrated Phonological Awareness: An intervention program for preschool children with Speech-language impairment. URL: https://www.canterbury.ac.nz/media/documents/education-and-health/gail-gillon---phonological-awareness-resources/programmes/preschool/01-Integrated-Phonological-Awareness-Manual-Sept-07.pdf [accessed 2019-11-05]

45. Bruns A, Jacobs J. Uses of Blogs. New York: Peter Lang; 2006.

46. de Villiers MR, Harpur PA. Design-based Research - The Educational Technology Variant of Design Research: Illustrated by the Design of an M-learning Environment. In: Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference. 2013 Presented at: SAICSIT'13; October 7-9, 2013; East London, South Africa p. 252-261. [doi: 10.1145/2513456.2513471]

47. Barab S, Squire K. Design-based research: putting a stake in the ground. J Learn Sci 2004 Jan;13(1):1-14. [doi: 10.1207/s15327809jls1301_1]

48. Read JC. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. Cogn Tech Work 2008;10(2):119-128. [doi: 10.1007/s10111-007-0069-9]

49. Phaser - A fast, fun and free open source HTML5 game. URL: http://phaser.io/ [accessed 2018-11-28] [WebCite Cache ID 74GQXmG9h]

50. Bootstrap. URL: https://getbootstrap.com/ [accessed 2018-11-28] [WebCite Cache ID 74GR7pYWR]

51. Apache Cordova. URL: https://cordova.apache.org/ [accessed 2018-11-28] [WebCite Cache ID 74GRSlXYr]

52. Gray D. Doing Research in the Real World. Second Edition. Los Angeles: SAGE; 2009.

53. Vairinhos M. ARTEFACTOS TANGÍVEIS E ADAPTÁVEIS NO AMBIENTE DOMÉSTICO. Portuguese: Universidade de Aveiro; 2014.

54. Abramson CM, Joslyn J, Rendle KA, Garrett SB, Dohan D. The promises of computational ethnography: improving transparency, replicability, and validity for realist approaches to ethnographic analysis. Ethnography 2018;19(2):254-284. [doi: 10.1177/1466138117725340]

Abbreviations

CSS: cascading style sheets
DBR: design-based research
FGTI: fishing game tangible interface
JSON: JavaScript Object Notation
KTAs: kindergarten teachers and assistants
LCD: liquid crystal display
RFID: radio frequency identification
SLT: speech and language therapist
SLTs: speech and language therapists
SSD: speech sound disorders
TUI: tangible user interface

XSL•FO
RenderX

Original Paper

A Serious Game–Derived Index for Detecting Children With Heterogeneous Developmental Disabilities: Randomized Controlled Trial

Changbae Bang[1], BS; Yelin Nam[1], BS; Eun Jae Ko[2], MD, PhD; Wooseong Lee[3], BS; Byungjae Kim[3], BS; Yejin Choi[3], BS; Yu Rang Park[4], PhD

[1]Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Rehabilitation Medicine, University of Ulsan College of Medicine, Ulsan, Republic of Korea

[3]DoBrain Co, Ltd, Seoul, Republic of Korea

[4]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics, Yonsei University College of Medicine
College of Medicine New Building 6F
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2228 2493
Email: yurangpark@yuhs.ac

Abstract

Background: Developmental disabilities are a set of heterogeneous delays or difficulties in one or more areas of neuropsychological development. Considering that childhood is an essential stage of brain development and developmental delays lead to personal or social burdens, the early detection of childhood developmental disabilities is important. However, early screening for developmental disabilities has been a challenge because of the fear of positive results, expensive tests, differences in diagnosis depending on examiners' abilities, and difficulty in diagnosis arising from the need for long-term follow-up observation.

Objective: This study aimed to assess the feasibility of using a serious game–derived index to identify heterogeneous developmental disabilities. This study also examines the correlation between the game-derived index and existing neuropsychological test results.

Methods: The randomized controlled trial involved 48 children with either normal development or developmental disabilities. In this clinical trial, we used 19 features (6 from the Korean-Wechsler Preschool and Primary Scale of Intelligence, 8 from the Psychoeducational Profile Revised, 2 from the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition, and 3 from the Pediatric Evaluation of Disability Inventory) from neuropsychological tests and 9 (7 game scores, path accuracy, and completion rate) from the serious game, DoBrain. The following analysis was conducted based on participants' baseline information and neuropsychological test and game-derived index data for one week: (1) we compared the baseline information between the normal development and developmental disabilities groups; (2) then we measured the correlation between the game-derived index and the neuropsychological test scores for each group; and (3) we built a classifier based on the game-derived index with a Gaussian process method and then compared the area under the curve (AUC) with a model based on neuropsychological test results.

Results: A total of 16 children (normal development=9; developmental disabilities=7) were analyzed after selection. Their developmental abilities were assessed before they started to play the serious games, and statistically significant differences were found in both groups. Specifically, the normal development group was more developed than the developmental disabilities group in terms of social function, gross motor function, full-scale IQ, and visual motor imitation, in that order. Similarly, the normal development group obtained a higher score on the game-derived index than the developmental disabilities group. In the correlation analysis between the game-derived index and the neuropsychological tests, the normal development group showed greater correlation with more variables than the developmental disabilities group. The game-derived index–based model had an AUC=0.9, a similar detection value as the neuropsychological test–based model's AUC=0.86.

XSL•FO
RenderX

Conclusions: A game-derived index based on serious games can detect children with heterogenous developmental disabilities. This suggests that serious games can be used as a potential screening tool for developmental disabilities.

Trial Registration: Clinical Research Information Service KCT0003247; https://cris.nih.go.kr/cris/en/search/search_result_st01 .jsp?seq=12365

Introduction

Developmental disabilities are among the most common diseases in children younger than 5 years old. Developmental disabilities are a set of heterogenous delays or difficulties in one or more developmental milestones, including learning, self-care, social interactions, and movement. The global number of children under 5 years old with developmental disabilities was 52.9 million in 2016, accounting for 13.3% of total years living with disability for these children. In 1990, 53.0 million children were living with developmental disabilities, indicating that there has not been much change since then [1]. In the United States, the prevalence of children diagnosed with developmental disabilities increased remarkably from 5.76% in 2014 to 6.99% in 2016 [2].

The brain is sensitive to stimulation during childhood, which is an essential stage of human development. It is the foundation for successive educational and vocational achievements, as well as society's human capital development [3]. However, children with developmental disabilities are at a higher risk of substandard educational accomplishment, health status, and social relationships. More specifically, children with developmental disabilities have heightened difficulties reading, spelling, and counting due to shortages of phonological short-term memory or central, executive-loaded, working memory [4]. Additionally, developmental disabilities lead to sedentary lifestyles and seven times greater reported substandard emotional support in adulthood than in adults without disabilities. Consecutive developmental disabilities are risk factors of chronic health conditions such as high blood pressure, cardiovascular disease, diabetes, and chronic pain [5].

If a child fails to achieve a certain milestone, it is extremely hard to recover that milestone later in life. Therefore, late identification of developmental disabilities in children may require schools and families to pay for expensive programs [6]. Additionally, an understanding of the subsequent characteristics of a child's developmental process requires a sensitive longitudinal examination by caregivers or specialists.

Given such a context, identifying developmental disabilities as early as possible has been deemed crucial. Multiple screening programs occurring in conjunction are recommended to detect early-stage developmental disabilities. For example, the American Academy of Pediatrics recommends frequent counseling, preventive care visits, and treatment visits for children [7]. Nevertheless, the early detection of developmental disabilities in children has not been performed adequately for many reasons, such as a lack of specific training for screening developmental disabilities, lack of time among specialists, fear of positive screen results, and failure to consider testing for developmental disabilities necessary [8]. In fact, even when early diagnosis and intervention programs are conducted in a timely manner, many instruments for diagnosis are administered in an environment distinct from an everyday setting. For example, children are directed to perform tasks in a laboratory setting or under the supervision of unfamiliar investigators.

To resolve these problems, approaches using serious games have been proposed for the early detection of developmental disabilities in children. Serious games are *games that do not have enjoyment, entertain[ment] or fun as their primary purpose* [9].

Serious games targeted for health care use have seen a surge in use since 2004 [10]. Health care–targeted serious games are intended for health checks, disease detection, or rehabilitation. Some studies have been conducted that investigated games created for detecting various disorders, such as Parkinson disease, Alzheimer disease, or early dementia [11-14]. Other serious games have been targeted towards detecting developmental disabilities in children. For example, Anzulewicz created a machine learning model that can identify children with autism [15]. Alchalabi also studied ways to detect children with attention deficit hyperactivity disorder (ADHD) using a serious game integrated with electroencephalogram signals [16]. Additionally, many researchers have targeted adult cognitive impairment detection [17-21].

However, thus far there have been no studies about detecting general and heterogeneous developmental disabilities in children using a serious game, despite how important it is to detect these developmental disabilities. Referring children to specialists for clinical evaluation is also the most essential part of an early diagnosis of developmental disabilities. Therefore, this study aimed to develop a classifier that can distinguish between children with heterogeneous developmental disabilities and those with normal development using a game-derived index based on a serious game for improving cognitive ability. In this study, we collected gameplay data of children from Do Brain, a smart device-based serious game intended for the cognitive enhancement of children.

Methods

Summary

This study was based on a single-blinded, parallel randomized controlled trial under the supervision of an independent data management team in the Asan Medical Center (Seoul, South

Korea) The study consists of an intervention and a control group, with an allocation ratio of 1:1. Participants were openly recruited and enrolled in clinical studies via face-to-face evaluation by physicians and inspectors of neuropsychological testing. Children with normal development were 5-6 years old and were confirmed to have developed normally via examinations and specialists. Children with developmental disabilities were 5-7 years old but had a cognitive age of 4-6 years old. The recruitment period for this study was from October 2018 to January 2019. If the child was assigned to a group after the face-to-face evaluation, intervention was provided for 6 months. Our study only used the results obtained in the first week of the intervention.

Participants were too young to give consent, so we received "Subject explanatory note and consent forms" from the participants' representatives (see Multimedia Appendix 1). The randomization, using a block size of four, was stratified depending on intervention and developmental disability. Randomization was done with the use of opaque, sealed envelopes. The statistician of the data management team generated the randomly allocated sequence with the use of the R program (R Core Team, Vienna, Austria). Physicians enrolled the patients and opened the envelope with the lowest available registration number within the appropriate stratum. Participants knew the group to which they belonged, but the physicians and the inspectors of the neuropsychological tests were blinded until the end of the trial.

The intervention of this study was based on a serious game and continued for 12 weeks for 40 minutes twice a week. The intervention was based on mobile serious gameplay at home.

There were no human interventions other than early, functional, game usage training. To encourage sustainable use, we sent a text message to the representatives of participants who had not used it once in the preceding week. All clinical trial data were documented, coded, and stored on a computer at Asan Medical Center in Seoul, Korea.

In the current study, we wanted to explore the possibility of differentiating children with developmental disabilities from the normal population by analyzing gameplay patterns. We used the results obtained in the first week of the intervention because, as the intervention progressed, the cognitive function of the participants changed. This is because the cognitive function of children changes much faster than that of adults, and the game Do Brain was intended to enhance the cognitive function of children. The intervention was designed for a particular time and frequency of play (40 minutes twice a week).

Initially, 106 participants were recruited. The medical staff screened 48 children who met the recruitment criteria. Therefore, the study group consisted of 48 children with either normal development or developmental disabilities (Figure 1). Twenty-two of the children underwent intervention for 12 weeks (experimental group) and the other 26 participants were the control group. The control group that did not undergo intervention and children who could not carry out level C of the serious game were excluded from this study. The baseline information and data of 16 eligible participants during the first week of the intervention were analyzed. This study was approved by the Institutional Review Board of Asan Medical Center (IRB #2018-0989; Seoul, South Korea).

Figure 1. Participant selection flow chart.
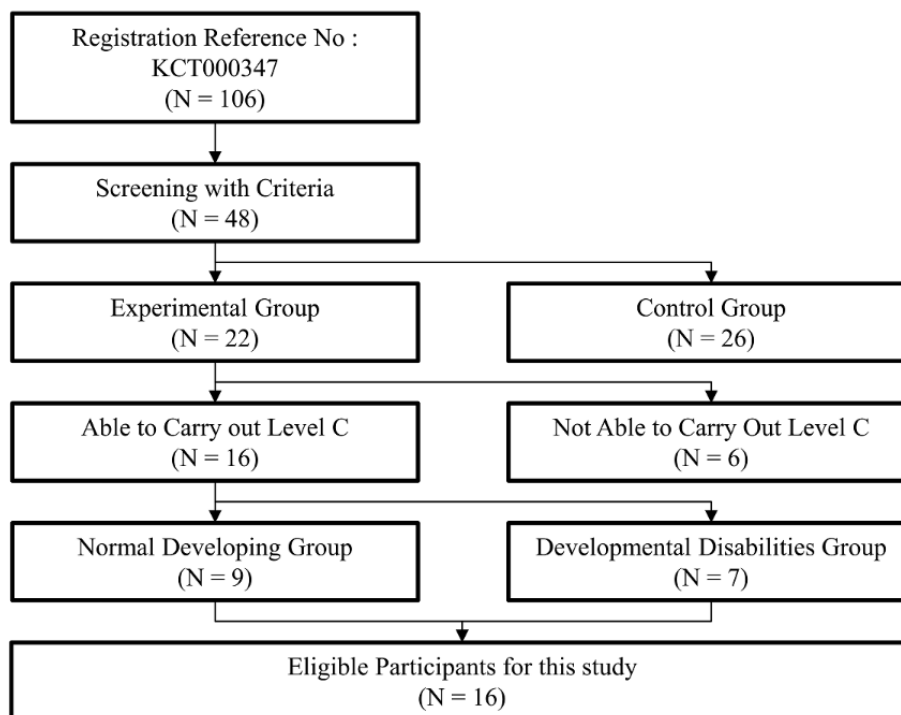
Neuropsychological Tests

A set of neuropsychological tests was conducted to assess children's developmental ability before the intervention period. This study employed the Korean-Wechsler Preschool and Primary Scale of Intelligence (K-WPPSI-IV) [22], a child's intelligence test that provides a comprehensive assessment of overall intellectual capabilities. The target age of the test is from 1 year and 6 months to 7 years and 7 months. This study also used the Full-Scale IQ, which provides five basic indicators of intellectual function of a particular cognitive domain, including Verbal Comprehension Index (VCI), Visual Spatial Index (VSI), Fluid Reasoning Index (FRI), Working Memory Index (WMI), and the Processing Speed Index (PSI). In addition, the Psychoeducational Profile Revised (PEP-R) [23], which us used to assess the treatment capacity of children between 1 and 7.5 years old with autism and related developmental disabilities, was used to plan treatment programs. Additionally, the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2) [24], which is administered in participants aged between 4 and 22 years, was used to check their motor development; it can also measure large and small muscle skills. The age range for the Pediatric Evaluation of Disability Inventory (PEDI) [25] is from 6 months to 7.5 years, and this test is used as a tool for evaluating independence in daily life through structured interviews with parents or caregivers.

Serious Game

Do Brain is a serious game based on an animated cartoon. The game is a smart device–based application certified for child suitability from the iOS App store and the Google Play Store. It consists of games that require simple touch inputs, such as one-point touch, drag and drop, and rub. The object of the game is to enhance primary cognitive capacity (attention, orientation, memory), higher-level thinking abilities (problem solving, reasoning, concept formation), and meta-processing abilities (executive function, self-awareness). The application was downloaded to the personal smart device of each participant's caregiver from the iOS App store or Google Play Store. The participants played the game in a natural environment, such as at home. A more detailed description is available at Do Brain homepage [26].

Game-Derived Index

In the program, there are sections that each contain 6-8 games. Each game of a section tests for each of the categories of development, which includes: Spatial Perception, Mathematical Thinking, Attention Memory, Logical Reasoning, Constructional Ability, Discernment, and Reaction (Figure 2). Each game has 1-3 stages. In the first week of the intervention participants played through nine sections, which consisted of 61 games and 128 stages. For the first week of the intervention we computed the game score, path accuracy, and completion rate.

The game score was calculated using both the duration and the incorrect answer count. Duration referred to the time it took a participant to answer correctly, and incorrect answer count to the number of wrong answers/attempts. The game score increased as the participant completed a certain stage within a shorter time and with fewer incorrect answers. Therefore, the higher the score a participant obtained, the better they played. Below is the equation for the game score:
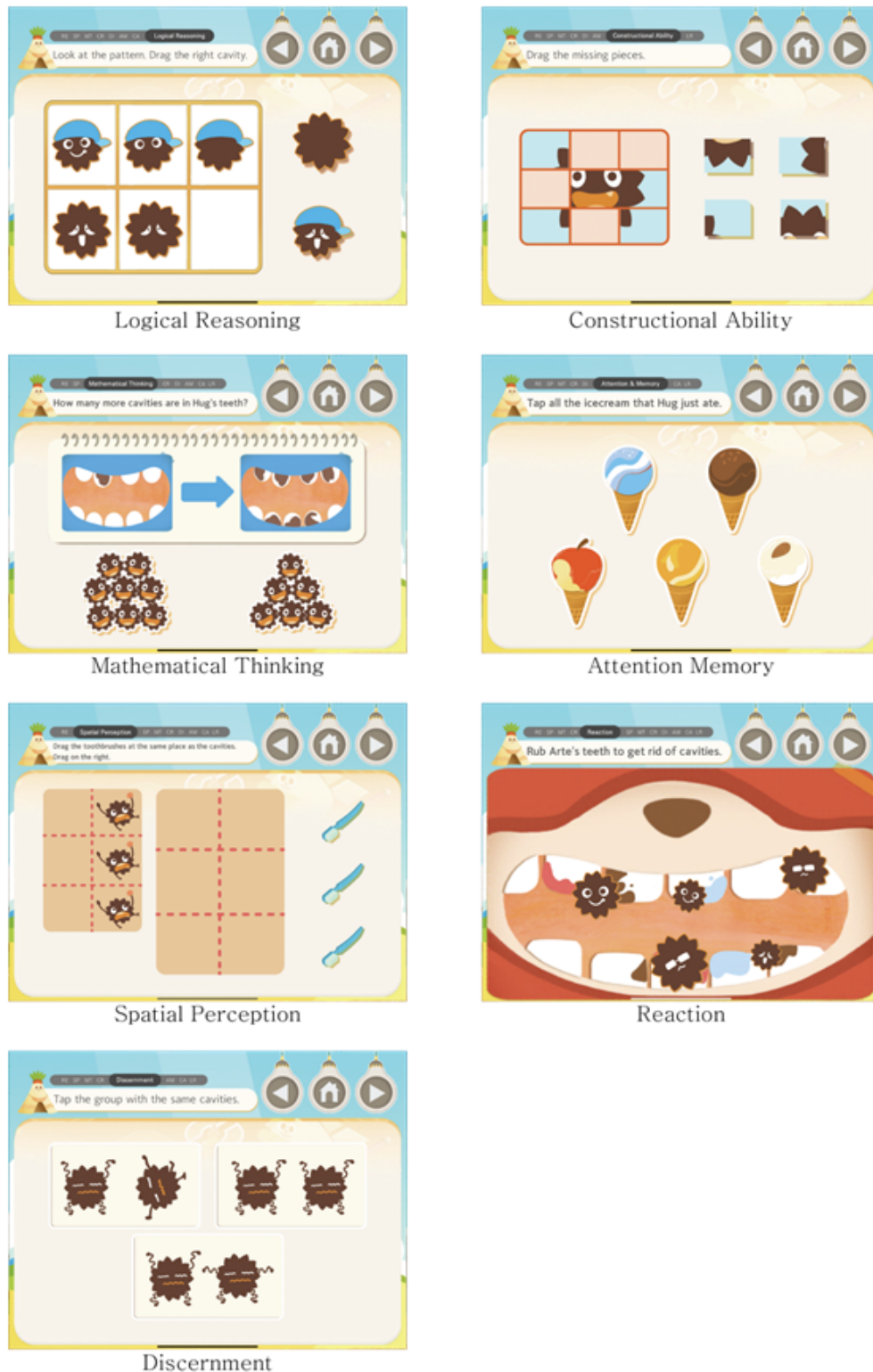

There were 16 drag-and-drop games in the 128 stages of the first week of the intervention, for which we computed path accuracy from both intended path and actual path. Intended path referred to the geometric distance between the start and end points and actual path to the distance of the finger's movement. The path accuracy value ranged from 0 to 1, with values closer to 1 indicating that a participant moved their finger more precisely. Below is the equation for path accuracy:

Meanwhile, completion rate was the ratio of the number of games played by each participant in a week divided by the quota for that week. It measured a participant's compliance.

Figure 2. Screenshots of the serious game by game categories.

Logical Reasoning

Constructional Ability

Mathematical Thinking

Attention Memory

Spatial Perception

Reaction

Discernment

Statistical Tests

In our study, we compared the baseline information between the normal development and developmental disabilities groups using the Mann-Whitney U test. We then measured the correlation between the game-derived index and a child's neuropsychological test scores for each group using Pearson's correlation method. Finally, we assessed the feasibility of using this serious game as a tool for detecting children with heterogeneous developmental disabilities. A classification model was built with a Gaussian process classifier, and model validation using leave-one-out cross validation (LOOCV) was conducted. LOOCV is a method for evaluating the prediction quality of a model built from a small dataset [27,28]. For the game-derived index–based model, we used nine features (7 game scores, path accuracy, and completion rate), and 19 features for the neuropsychological test–based model (6 K-WPPSI, 8 PEP-R, 2 BOT-2, 3 PEDI). In this study the

multi-variate Gaussian process classifier of Rasmussen and Williams was used [29]. The classifier is based on Laplace approximation and makes predictions based on finite combinations of all random variables that have multivariate normal distributions. The receiver operating characteristic (ROC) and precision-recall (PR) curves were drawn from the validation results, and then the area under the curve (AUC) was calculated for each curve. Sensitivity, specificity, precision, true positive rate, and true negative rate were also measured. We set a significance level of 0.05. Data were processed and analyzed using R version 3.5.0 (R Core Team, Vienna, Austria), and Python 3.6 (Python Software Foundation, Wilmington, Delaware, United States; including the Pandas 0.22.0, NumPy 1.14.3, and Jupyter 1.0.0 packages).

Results

Participant Characteristics

Baseline Information

The normal development and developmental disabilities groups were comprised of nine and seven children, respectively. Chronological age was not significantly different between the two groups, but sex was significantly different in that the proportion of males in the developmental disabilities group was higher than in the normal development group. The K-WPPSI-IV test results, which measured a child's cognitive ability, showed that full scale IQ was significantly different between the normal development and developmental disabilities groups (P=.008). Moreover, the normal development group surpassed the developmental disabilities group in all subcategories of K-WPPSI-IV. In particular, the normal development and developmental disabilities groups showed statistically significant differences in VCI (P=.01), VSI (P=.01), and FRI (P=.01; Table 1).

The developmental ages of the normal development and developmental disabilities groups were significantly different (P=.007). For all subcategories of PEP-R, the normal development group outperformed the developmental disabilities group. Specifically, statistically significant differences in Cognitive Verbal (P=.01), Cognitive Motion (P=.01), Fine Motor (P=.01), and Imitation (P=.03) categories were observed. Moreover, Visual Motor Imitation (P=.009) and Gross Motor (P=.005) showed more significant differences than the other categories (Table 1). For the data presented in Table 1, the Mann-Whitney U test was used as the statistical significance test.

Table 1. Comparison of the baseline information and neuropsychological tests of the normal development and developmental disability groups.

	Normal development group (n=9)	Developmental disability group (n=7)	*P* value
Sex, n (%)			
Male	4 (44.4)	6 (85.7)	__a
Female	5 (55.6)	1 (14.3)	—
Chronological age, mean (SD)	71.4 (6.20)	70.1 (8.01)	.35
K-WPPSI-IV[b], mean (SD)			
Full scale IQ	110.7 (18.3)	86.7 (13.7)	.008
Verbal comprehension index	113.2 (16.7)	89.1 (15.7)	.02
Visual spatial index	109.4 (13.8)	89.1 (19.0)	.01
Fluid reasoning index	104.6 (11.6)	89.0 (13.7)	.01
Working memory index	118.4 (16.8)	106.1 (15.5)	.14
Processing speed index	99.3 (16.2)	88.1 (14.5)	.14
PEP-R[c], mean (SD)			
Developmental age	68.7 (5.0)	62.0 (5.4)	.007
Cognitive verbal	73.6 (4.3)	64.4 (8.5)	.01
Cognitive motion	70.0 (4.8)	63.4 (6.1)	.01
Visual motor imitation	67.2 (6.2)	58.6 (5.4)	.009
Gross motor	59.6 (0.9)	52.4 (11.7)	.005
Fine motor	63.3 (4.2)	56.6 (5.4)	.01
Perception	59.4 (2.9)	58.5 (2.9)	.07
Imitation	63.1 (3.5)	60.6 (3.5)	.03
BOT-2[d], mean (SD)			
Fine motor	46.4 (19.9)	29.1 (13.8)	.05
Manual coordination	30.2 (9.0)	22.3 (8.8)	.08
PEDI[e], mean (SD)			
Self-care	70.5 (1.5)	66.0 (5.6)	.10
Movement	57.0 (3.7)	54.4 (5.6)	.20
Social function	62.7 (2.5)	56.7 (4.6)	.005

[a]Not applicable.

[b]K-WPPSI-IV: Korean-Wechsler Preschool and Primary Scale of Intelligence.

[c]PEP-R: Psychoeducational Profile Revised.

[d]BOT-2: Bruininks-Oseretsky Test of Motor Proficiency, Second Edition.

[e]PEDI: Pediatric Evaluation of Disability Inventory.

Game-Derived Index

Overall, the normal development group showed better game scores than the developmental disabilities group on the game-derived index (Table 2). Logical Reasoning (*P*=.03), Constructional Ability (*P*=.04), Mathematical Thinking (*P*=.01), and Attention Memory (*P*=.04) were statistically significant in the two groups. In all subcategories of the game-derived index, the developmental disabilities group's standard deviation was larger than in the normal development group. For the data presented in Table 2, the Mann-Whitney *U* test was used as the statistical significance test.

Table 2. Comparison of the game-derived index between the normal development and developmental disability groups.

Variable	Normal development group (n=9)	Developmental disability group (n=7)	P value
Game score, mean (SD)			
Logical reasoning	8.28 (0.66)	7.16 (1.16)	.03
Constructional ability	7.82 (0.69)	7.04 (1.01)	.04
Mathematical thinking	9.12 (0.54)	8.13 (0.88)	.01
Attention memory	9.20 (0.68)	8.50 (0.98)	.04
Spatial perception	8.67 (0.58)	7.76 (0.97)	.06
Reaction	8.31 (0.78)	7.89 (0.30)	.05
Discernment	8.71 (0.55)	8.07 (1.07)	.19
Path accuracy, mean (SD)	0.795 (0.089)	0.750 (0.072)	.17
Completion rate, mean (SD)	0.991 (0.026)	0.910 (0.203)	.10

Correlations Between the Neuropsychological Tests and Game-Derived Index

Correlations between the neuropsychological tests and game-derived index results were computed for each group using Pearson's correlation method (Figure 3). Overall, the game-derived index and neuropsychological tests showed higher correlation values for the normal development group than the developmental disabilities group. Logical reassignment and completion rate in the game-derived index showed a high correlation with neuropsychological tests. In the developmental disabilities group, the constructional ability of the game-derived index showed the highest correlation with VSI (0.92).

Game-Derived Index– and Neuropsychological Test–Based Classifiers

As shown in Figures 4 and 5, the game-derived index–based, classifier has a similar AUC to the neuropsychological test–based classifier. Sensitivity and specificity showed similar patterns. The game-derived index classifier had a sensitivity of 0.714 and a specificity of 0.778, whereas the neuropsychological test–based classifier had a sensitivity of 0.857 and a specificity of 0.778.

No serious adverse events or side effects were observed in the normal development and developmental disabilities groups.

Figure 3. Correlation matrix for the game-derived index and neuropsychological tests of the normal development and developmental disability groups. K-WPPSI-IV: Korean-Wechsler Preschool and Primary Scale of Intelligence; PEP-R: Psychoeducational Profile Revised; BOT-2: Bruininks-Oseretsky Test of Motor Proficiency, Second Edition; PEDI: Pediatric Evaluation of Disability Inventory.
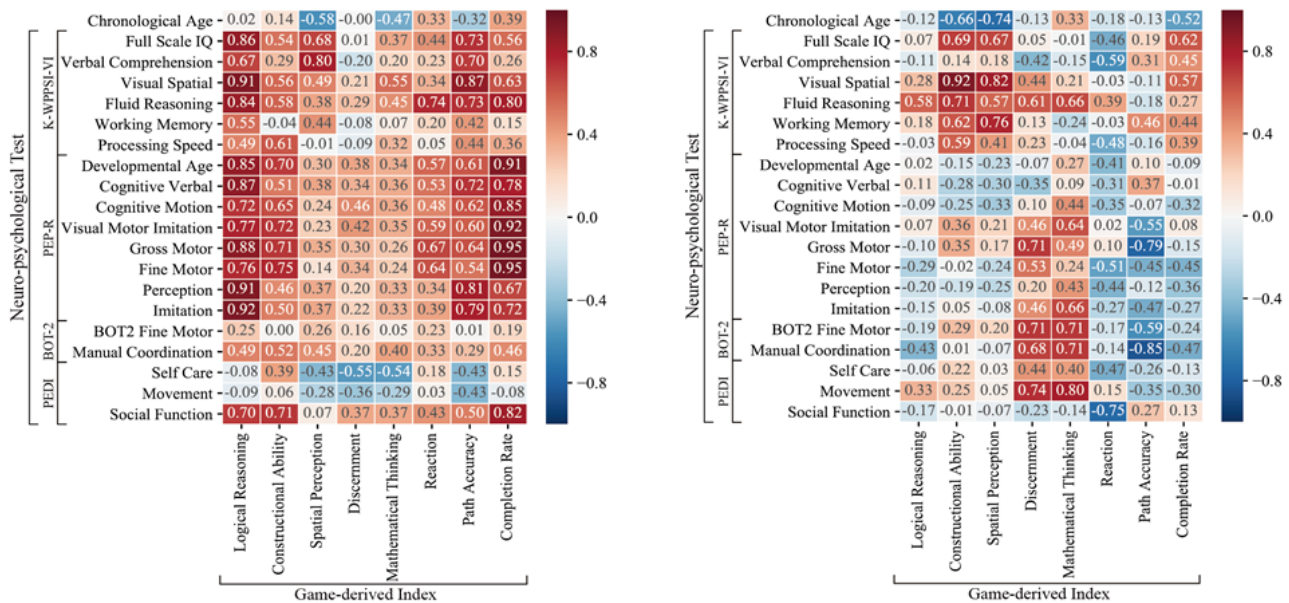
(a) Normal Developing Group

(B) Developmental Disability Group

Figure 4. The receiver operating characteristic and precision-recall curves for leave-one-out cross validation of the model based on the game-derived index from the serious game. ROC: receiver operating characteristic; DD: developmental disabilities; ND: normal development; PR: precision recall; GDI: game-derived index.
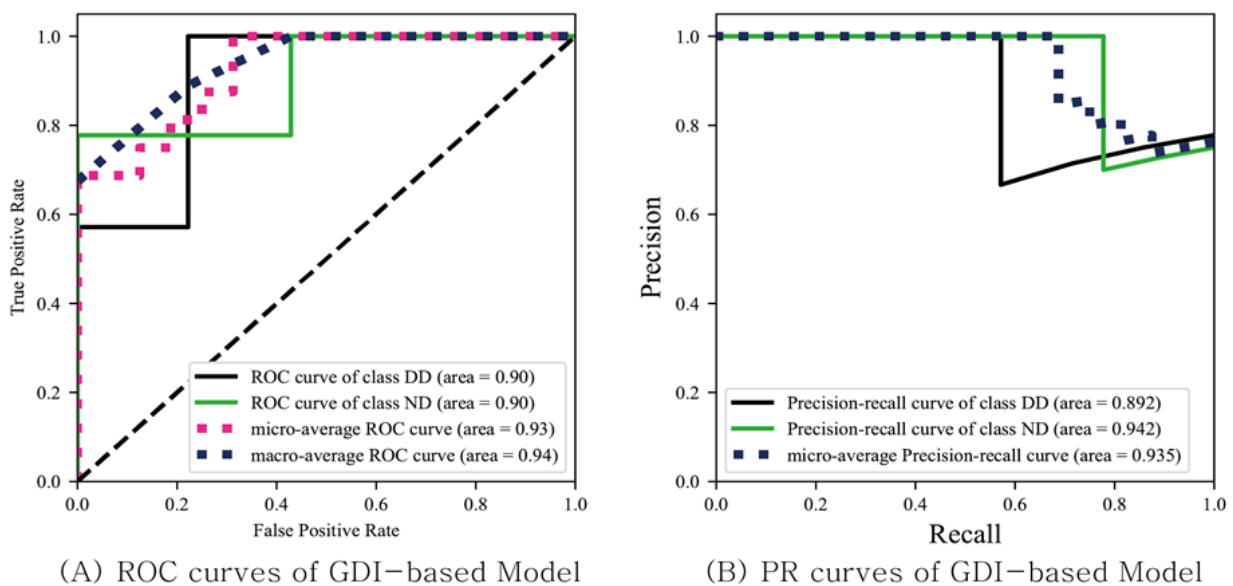

(A) ROC curves of GDI-based Model

(B) PR curves of GDI-based Model

Figure 5. The receiver operating characteristic and precision-recall curves for leave-one-out cross-validation of the model based on the neuropsychological tests. ROC: receiver operating characteristic; DD: developmental disabilities; ND: normal development; PR: precision recall.

(A) ROC curves of neuropsychological tests based Model

(B) PR curves of neuropsychological tests based Model
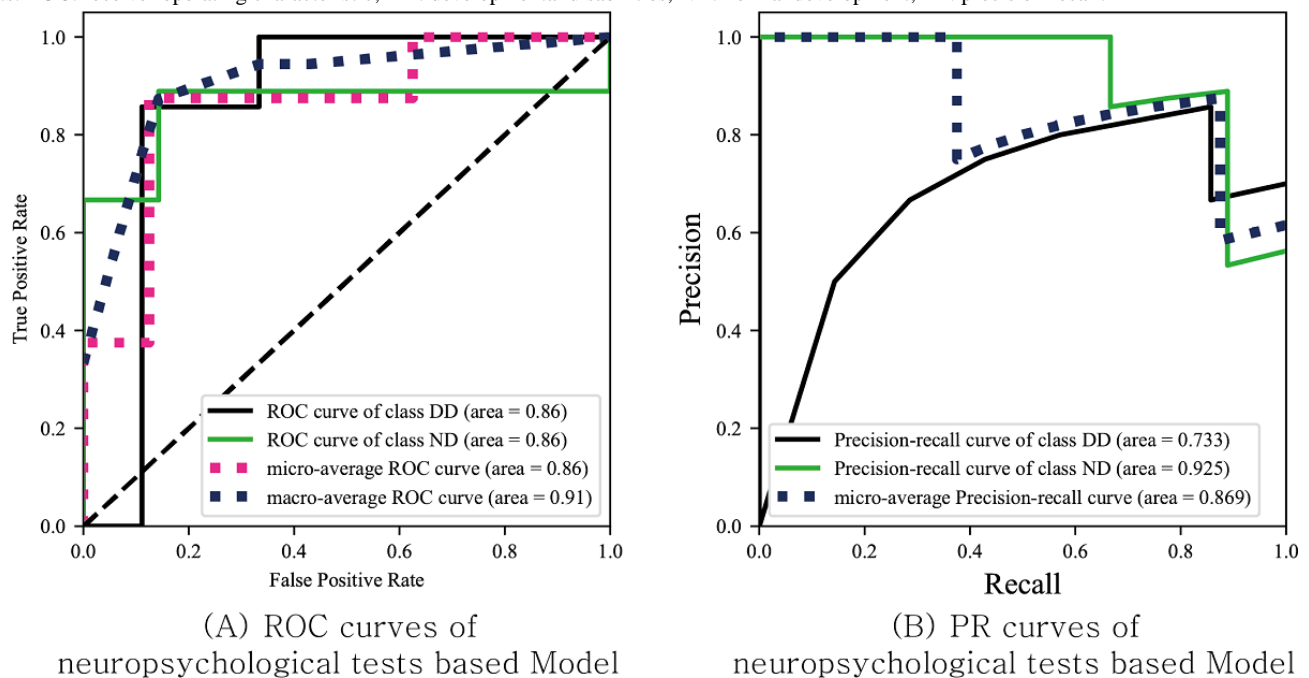
Discussion

Principal Results

We built a classifier from the game-derived index that could distinguish children with developmental disabilities from those with normal development. In contrast to some models that can detect a specific type of disease, such as autism or ADHD, our classifier is the first model to detect children with heterogeneous types of developmental disabilities. This suggests that serious games have the potential for screening the general developmental disability status of children. Considering that

the positive diagnosis of developmental disabilities should be conducted through a specialist's exam, our classifier that detects general and heterogeneous developmental disabilities could be used to refer children with developmental disabilities to specialists for more specific and expert exams.

Additionally, we compared the game-derived index with the results of the neuropsychological tests to determine the characteristics of the normal development group that helped them obtain a better game-derived index than the developmental disabilities group. As many serious games that assess the cognitive function of humans lack psychometric analysis and

comparisons with standardized studies [30], the comparison between the game-derived index and the neuropsychological tests in this study provides a more comprehensive interpretation of the serious gameplay of children.

Differences in Game-Derived Index Between Normal Development and Developmental Disabilities Groups

Game score, path accuracy, and completion rate were computed to assess participants' gameplay. Although some features of the game-derived index lacked statistical significance, the overall pattern showed that the normal development group played better than the developmental disabilities group.

Game score represents how fast and correctly a participant solved a problem. The serious game used in this study is based on a cognitive counseling program for children. As the normal development group showed higher intelligence scores on the neuropsychological tests, this means that they answered correctly with fewer attempts and in less time. For the normal development group, many game-derived index features showed a positive linear correlation with neuropsychological tests. More specifically, Logical Reasoning was correlated with Full Scale IQ and Developmental Age with coefficients of 0.86 and 0.85, respectively.

By contrast, most game-derived index features of the developmental disabilities group did not show linear correlations or had negative correlations. These results can be attributed to the following reasons. First, the developmental disabilities were heterogenous in the developmental disabilities group. Also, the developmental disabilities group had a small, restricted sample. Moreover, some studies indicate that cognitive assessment is difficult for patients suffering from neuropsychological disorders [31-33]. More specifically, when converting raw scores to standardized scores, performance variations that need to be detected often become obscure due to the flooring effect [34,35]. The flooring effect refers to a phenomenon in which the measuring tool cannot discriminate among those who belong to the lower level of the characteristic to be measured [36]. The flooring effect may occur when the score range itself is limited, or when inspection is too difficult. As standardized tests have such limitations, recent trends in intellectual disability research emphasizes narrating the cognitive signatures of conditions throughout their lifetime rather than depending on test scores [31]. Although many game-derived index features did not have positive correlations, constructional ability, spatial perception, and completion rate had correlation coefficients of 0.69, 0.67, and 0.62, respectively.

Path accuracy indicates fine motor function and visuospatial ability. According to Vatavu, children who are better developed in visuospatial processing have a high path accuracy score [37]. In our study, the normal development group's path accuracy score was higher than the developmental disabilities group, but the difference was not statistically significant. Path accuracy scores were highly correlated with the VSI of the K-WPPSI in the normal development group (r=0.87), which shows a similar pattern to Vatavu's findings. These findings indicate that children with better VSI scores are better at interpreting geometrical relationships. As a result, they understand drag-and-drop games better and can move their finger in a more intentional path. However, the normal development group's path accuracy did not correlate well with Fine Motor index of PEP-R and BOT-2, mainly because testing of PEP-R or BOT-2 differs from a drag-and-drop game. Moreover, the developmental disabilities group's path accuracy had negative correlation with the neuropsychological tests' VSI and Fine Motor index. This negative correlation can be attributed to the developmental disabilities group's heterogeneity and small sample size.

Completion rate is a measure of participant compliance. In this serious game, participants may skip a stage when they find it difficult. As can be observed in the game score results, the normal development group outperformed the developmental disabilities group in gameplay. Additionally, the normal development group's mean completion rate was higher than the developmental disabilities group's. Notably, the standard deviation differed greatly between the normal development (0.026) and developmental disabilities (0.203) groups. The large standard deviation of the developmental disabilities group's completion rate is due to some developmental disabilities participants skipping many more stages than the others.

Classification Model

The model built from the game-derived index performed adequately compared with that built from neuropsychological tests, showing the discriminating power of the game-derived index. This result suggests that serious games have the potential for detecting children with developmental disabilities. As children with heterogeneous developmental disabilities show complicated characteristics, characterizing them with high sensitivity is difficult. In contrast to models that detect specific developmental disorders developed in previous studies, the model built in this study is the first to detect general developmental disabilities.

Comparison With Previous Studies

Some studies have been conducted regarding developmental delay detection with serious games. Previous works utilized more devices, required more resources, or could detect only specific diseases.

Alchalabi's study that detected ADHD patients with an Electroencephalogram (EEG)-based serious game [16] required an EEG reader and a special gaming setting. However, our study only utilizes smart devices and is done in the home setting. Like our study's use of neuropsychological tests for clinical diagnosis, Alchalabi's study was also meaningful in that it used EEG with some diagnostic value.

Elhady build a speech disability detection game with limited speech resources [38]. Our study uses touch input from a smart device while it records the voice. However, they targeted a very specific speech disability population suffering from the phonemes /s/ and /r/. There is a need to cover broader speech disabilities, as there exist heterogenous speech disabilities [39].

Garcia and Ruiz created a children's psychomotor delay screening process called Ubiquitous Detection Ecosystem to Care and Early Stimulation for Children with Developmental Disorders (EDUCARE) [40] based on a smart toy. In

EDUCARE, the video that children played on the toy had been analyzed by developmental experts. However, EDUCARE only assessed motor function and depended on lengthy screening by experts. Our study overcomes this by assessing more developmental areas and immediately giving numerical developmental indices to caregivers. In a different respect, EDUCARE utilized methods more related to everyday objects, such as stackable cubes, than to our game-based approach.

Limitations

Our study population is insufficient for generalizability of our findings to the general population. However, our findings show the feasibility of using this serious game to screen for general and heterogeneous developmental disabilities. As children continue to develop in various areas in the earlier stages of developmental disabilities, a definite diagnosis of the disease is difficult. Our model is useful because of the flexibility and heterogeneous characteristics of developmental disabilities. In the future, we will widen our study population to enhance the generalizability of our findings.

In a clinical situation, a neuropsychological test is essential for the diagnosis of developmental disabilities. Therefore, the estimation of neuropsychological test results with a game-derived index could provide caregivers and specialists with more information. However, our study did not build a model that estimates neuropsychological test results from the game-derived index because of the small sample size. Future studies must build a neuropsychological test estimation model more complex than our simple linear correlation model.

Conclusion

The game-derived index patterns observed in serious gameplay in the normal development and developmental disabilities groups were different. In particular, the neuropsychological tests and game-derived index have significant differences in the correlations between groups. This result is a potential indicator that game-derived index can distinguish developmental disabilities from normal development, and the model based on it showed similar performance to neuropsychological tests that constitute conventional developmental ability tests. This suggests that serious games can be used as a potential screening tool for developmental disabilities.

Conflicts of Interest

WL, BK, and YC are employees of Do Brain Co, Ltd.

Multimedia Appendix 1
Informed consent for the randomized controlled trial.
[DOCX File , 55 KB - games_v7i4e14924_app1.docx]

Multimedia Appendix 2
CONSORT‐EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3666 KB - games_v7i4e14924_app2.pdf]

References

1. Global Research on Developmental Disabilities Collaborators. Developmental disabilities among children younger than 5 years in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Glob Health 2018 Oct;6(10):e1100-e1121 [FREE Full text] [doi: 10.1016/S2214-109X(18)30309-7] [Medline: 30172774]
2. Zablotsky B, Black LI, Blumberg SJ. Estimated Prevalence of Children With Diagnosed Developmental Disabilities in the United States, 2014-2016. NCHS Data Brief 2017 Nov(291):1-8 [FREE Full text] [Medline: 29235982]
3. Shonkoff JP, Phillips DA. From Neurons to Neighborhoods: The Science of Early Childhood Development. Washington, DC: National Academies Press; 2000.
4. Henry L, Winfield J. Working memory and educational achievement in children with intellectual disabilities. J Intellect Disabil Res 2010 Apr;54(4):354-365. [doi: 10.1111/j.1365-2788.2010.01264.x] [Medline: 20236191]
5. Havercamp SM, Scandlin D, Roth M. Health disparities among adults with developmental disabilities, adults with other disabilities, and adults not reporting disability in North Carolina. Public Health Rep 2004;119(4):418-426 [FREE Full text] [doi: 10.1016/j.phr.2004.05.006] [Medline: 15219799]
6. The importance of caregiver-child interactions for the survival and healthy development of young children. In: Developmental difficulties in early childhood. Boston, US: World Health Organization; 2012.

XSL·FO

RenderX

7. Committee On Practice and Ambulatory Medicine, Bright Futures Periodicity Schedule Workgroup. 2019 Recommendations for Preventive Pediatric Health Care. Pediatrics 2019 Mar 25;143(3):e20183971. [doi: 10.1542/peds.2018-3971] [Medline: 30804073]

8. Pinto-Martin JA, Dunkle M, Earls M, Fliedner D, Landes C. Developmental Stages of Developmental Screening: Steps to Implementation of a Successful Program. Am J Public Health 2005 Nov;95(11):1928-1932. [doi: 10.2105/ajph.2004.052167]

9. Michael D, Chen S. Serious Games: Games that Educate, Train and Inform. Boston: Thomson Course Technology; 2006.

10. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertainment Computing 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]

11. van der Meulen E, Cidota MA, Lukosch SG, Bank PJM, van der Helm AJC, Visch VT. A Haptic Serious Augmented Reality Game for Motor Assessment of Parkinson's Disease Patients. 2016 Presented at: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct); 19-23 Sept. 2016; Merida p. 104.

12. Vallejo V, Wyss P, Rampa L, Mitache AV, Müri RM, Mosimann UP, et al. Evaluation of a novel Serious Game based assessment tool for patients with Alzheimer's disease. PLoS One 2017;12(5):e0175999 [FREE Full text] [doi: 10.1371/journal.pone.0175999] [Medline: 28472049]

13. Tarnanas I, Laskaris N, Tsolaki M, Muri R, Nef T, Mosimann UP. On the comparison of a novel serious game and electroencephalography biomarkers for early dementia screening. Adv Exp Med Biol 2015;821:63-77. [doi: 10.1007/978-3-319-08939-3_11] [Medline: 25416111]

14. Zucchella C, Sinforiani E, Tassorelli C, Cavallini E, Tost-Pardell D, Grau S, et al. Serious games for screening pre-dementia conditions: from virtuality to reality? A pilot project. Funct Neurol 2014;29(3):153-158 [FREE Full text] [Medline: 25473734]

15. Anzulewicz A, Sobota K, Delafield-Butt JT. Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. Sci Rep 2016 Aug 24;6(1):31107 [FREE Full text] [doi: 10.1038/srep31107] [Medline: 27553971]

16. Alchalabi AE, Shirmohammadi S, Eddin AN, Elsharnouby M. FOCUS: Detecting ADHD Patients by an EEG-Based Serious Game. IEEE Trans. Instrum. Meas 2018 Jul;67(7):1512-1520. [doi: 10.1109/tim.2018.2838158]

17. Valladares-Rodriguez S, Perez-Rodriguez R, Facal D, Fernandez-Iglesias MJ, Anido-Rifon L, Mouriño-Garcia M. Design process and preliminary psychometric study of a video game to detect cognitive impairment in senior adults. PeerJ 2017;5:e3508 [FREE Full text] [doi: 10.7717/peerj.3508] [Medline: 28674661]

18. Vallejo V, Wyss P, Chesham A, Mitache AV, Müri RM, Mosimann UP, et al. Evaluation of a new serious game based multitasking assessment tool for cognition and activities of daily living: Comparison with a real cooking task. Computers in Human Behavior 2017 May;70:500-506. [doi: 10.1016/j.chb.2017.01.021]

19. Tong T, Chignell M, Tierney MC, Lee J. A Serious Game for Clinical Assessment of Cognitive Status: Validation Study. JMIR Serious Games 2016 May 27;4(1):e7 [FREE Full text] [doi: 10.2196/games.5006] [Medline: 27234145]

20. Byun S, Park C. Serious Game for Cognitive Testing of Elderly. In: HCI 2011: HCI International 2011–Posters' Extended Abstracts. Berlin: Springer; 2019:354-357.

21. Boletsis C, McCallum S. Smartkuber: A Serious Game for Cognitive Health Screening of Elderly Players. Games Health J 2016 Aug;5(4):241-251. [doi: 10.1089/g4h.2015.0107] [Medline: 27192473]

22. Park H, Lee K. Intellectual Characteristics of Korean Children: Analyses of K-WPPSI-IV Performance. Korean J Child Stud 2016 Dec 31;37(6):157-168. [doi: 10.5723/kjcs.2016.37.6.157]

23. Schopler E. Psychoeducational Profile Revised (PEP-R). In: Encyclopedia of Autism Spectrum Disorders: 5 Volume Set. New York: Springer; 2019.

24. Deitz JC, Kartin D, Kopp K. Review of the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2). Phys Occup Ther Pediatr 2007;27(4):87-102. [Medline: 18032151]

25. Haley SM, Coster WJ, Ludlow LH, Haltiwanger JT, Andrellos PJ. Pediatric evaluation of disability inventory (PEDI): Development, standardization and administration manual. New York: The Psychological Corporation USA; 1997.

26. DOBY-Early Start in Math. 2019. URL: https://dobrain.co/ [accessed 2019-10-09]

27. Chen X, Huang L, Xie D, Zhao Q. EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. Cell Death Dis 2018 Jan 05;9(1):3 [FREE Full text] [doi: 10.1038/s41419-017-0003-x] [Medline: 29305594]

28. You Z, Huang Z, Zhu Z, Yan G, Li Z, Wen Z, et al. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. PLoS Comput Biol 2017 Mar 24;13(3):e1005455 [FREE Full text] [doi: 10.1371/journal.pcbi.1005455] [Medline: 28339468]

29. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge: MIT Press; 2005.

30. Valladares-Rodríguez S, Pérez-Rodríguez R, Anido-Rifón L, Fernández-Iglesias M. Trends on the application of serious games to neuropsychological evaluation: A scoping review. J Biomed Inform 2016 Dec;64:296-319 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.019] [Medline: 27815228]

31. Kogan CS, Boutet I, Cornish K, Graham GE, Berry-Kravis E, Drouin A, et al. A comparative neuropsychological test battery differentiates cognitive signatures of Fragile X and Down syndrome. J Intellect Disabil Res 2009 Feb;53(2):125-142. [doi: 10.1111/j.1365-2788.2008.01135.x] [Medline: 19054268]

32.   Hessl D, Nguyen DV, Green C, Chavez A, Tassone F, Hagerman RJ, et al. A solution to limitations of cognitive testing in children with intellectual disabilities: the case of fragile X syndrome. J Neurodev Disord 2009 Mar 22;1(1):33-45 [FREE Full text] [doi: 10.1007/s11689-008-9001-8] [Medline: 19865612]

33.   Ozonoff S. Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. Neuropsychology 1995;9(4):491-500. [doi: 10.1037//0894-4105.9.4.491]

34.   Groth-marnat G. Handbook of Psychological Assessment. Hoboken: Wiley; 2009.

35.   Sternberg JR, Kaufman BS. The Cambridge Handbook of Intelligence. Cambridge: Cambridge University Press; 2019.

36.   Everitt BS, Skrondal A. The Cambridge Dictionary of Statistics. Cambridge: Cambridge University Press; 2010.

37.   Vatavu R, Cramariuc G, Schipor DM. Touch interaction for children aged 3 to 6 years: Experimental findings and relationship to motor skills. International Journal of Human-Computer Studies 2015 Feb;74:54-76. [doi: 10.1016/j.ijhcs.2014.10.007]

38.   Reem E, Mohamed E, Injy H, Slim A. A Game with a Purpose for Automatic Detection of Children's Speech Disabilities Using Limited Speech Resources. 2018 Jan 05 Presented at: International Conference on Arabic Language Processing; 11-12 October 2017; Fez, Morocco p. 89. [doi: 10.1007/978-3-319-73500-9_6]

39.   Verdolini K, Rosen CA, Branski RC. Classification Manual For Voice Disorders-I. Abingdon-on-Thames: Routledge; 2019.

40.   Gutiérrez García MA, Martín Ruiz ML, Rivera D, Vadillo L, Valero Duboy MA. A Smart Toy to Enhance the Decision-Making Process at Children's Psychomotor Delay Screenings: A Pilot Study. J Med Internet Res 2017 May 19;19(5):e171 [FREE Full text] [doi: 10.2196/jmir.7533] [Medline: 28526666]

Abbreviations

ADHD: attention deficit hyperactivity disorder
AUC: area under the curve
BOT-2: Bruininks-Oseretsky Test of Motor Proficiency, Second Edition
EDUCARE: Ubiquitous Detection Ecosystem to Care and Early Stimulation for Children with Developmental Disorders
EEG: electroencephalogram
FRI: Fluid Reasoning Index
K-WPPSI-IV: Korean-Wechsler Preschool and Primary Scale of Intelligence
LOOCV: leave-one-out cross validation
PEDI: Pediatric Evaluation of Disability Inventory
PEP-R: Psychoeducational Profile Revised
PR: precision recall
PSI: Processing Speed Index
ROC: receiver operating characteristic
VCI: Verbal Comprehension Index
VSI: Visual Spatial Index
WMI: Working Memory Index

XSL•FO
RenderX