Objective Method for User Experience Evaluation of Children with SLD on HapHop-Physio

Carolina Rico Olarte

Master's Thesis in Telematics Engineering

Supervisors:
Dr. Diego Mauricio López Gutiérrez, PhD
Department of Telematics, Universidad del Cauca
Dr. Sara Kepplinger, PhD
Fraunhofer Institute of Digital Media and Technology

Universidad Del Cauca
Faculty of Electronics and Telecommunications Engineering
Department of Telematics
eHealth
Popayán, November 2018

Carolina Rico Olarte

Objective Method for User Experience Evaluation of Children with SLD on HapHop-Physio

Dissertation submitted to the
Faculty of Electronics and Telecommunications Engineering
of the Universidad del Cauca, Colombia
for granting the academic degree of

Magíster en:
Ingeniería Telemática

Supervisors:
Dr. Diego Mauricio López Gutiérrez, PhD
Department of Telematics, Universidad del Cauca
Dr, Sara Kepplinger, PhD
Fraunhofer Institute of Digital Media and Technology

Popayán
2018

Universidad del Cauca

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Programas de Posgrado

FORMATO I
ACTA DE SUSTENTACIÓN DE
TRABAJO DE GRADO DE MAESTRÍA/TESIS DOCTORAL

Los jurados de:

TRABAJO DE GRADO DE MAESTRÍA (X) TESIS DOCTORAL ( )

Titulado:
Objective Method for User Experience Evaluation of Children with SLD on HapHop-Physio

Bajo la dirección de: Dr. Diego Mauricio López and Dr. Dara Kepplinger

HACEN CONSTAR:

Que siendo las 1:00 pm del día 7 del mes de noviembre de 2018, la estudiante:

Carolina Rico Olarte

realizó la Sustentación del Trabajo de Grado de Maestría/Tesis Doctoral, obteniendo la calificación de:

APROBADO (x) NO APROBADO ( ).

Para constancia, se firma en Popayán, a los 7 días del mes de noviembre de 2018.

JURADO 1 (o COORD. DEL PROGRAMA*):

Helmuth Trefftz

Nombre: Dr. Helmuth Trefftz

JURADO 2 (Doctorado):

Nombre:

JURADO COORDINADOR:

Luz Santamaría

Nombre: Mag. Luz Santamaría Granados

(*) Firma el Coordinador del Programa con autorización del(los) Jurado(s):

Para Huber y Lilián

Acknowledgements

First of all, I would like to thank God for making all this possible.

I must thank to doctor Diego Mauricio for making me a better researcher with their tutoring and advice, and to doctor Sara, for always keeping up our backs.

Also, I want to thank Universidad del Cauca and InnovAcción Cauca for supporting this research, the GIT group and the Fraunhofer IDMT for their scientific and personal support.

Many thanks to Nathalia Narvaez, the specialist that helped us in every aspect of the study with the children.

Finally, thanks to my family and friends for making this process more pleasant.

Structured abstract

Background: The user experience aims to determine, in a broad sense, the state of the interactions of a user with a product, system or service. Particularly for the services that the area of the Information and Communication Technologies provides to the health sector, the evaluation of the user experience becomes an important milestone in the design and implementation of this type of services, to maintain the motivation of the user when using them. However, the evaluation of the experience can be intrusive for the user and the results are biased by the subjectivity of the user and the evaluator. Evaluation methods based on psychophysiology gather the relationships that may exist between what the user feels, thinks and behaves, and what can be measured in their body in a non-intrusive way, to achieve objective results in the evaluations of the user experience.

Objectives: The main objective of this Master's thesis is the definition of a method for the objective evaluation of the user experience for children with specific learning disorders. To define the method, it is important to frame it in a conceptual framework identifying the psychophysiological measures used in the objective evaluation of applications in the eHealth area. The instantiation of the defined method leads to obtaining the measurement procedure that defines the relationship between a physiological signal and a psychological event. The method is evaluated within the HapHop-Fisio scenario to determine the user experience.

Methods: This thesis was divided into three phases to achieve the objective. First, the state of the art survey was carried out based on the methodologies of systematic mapping and systematic review to obtain the conceptual basis of this research work. Second, by means of a methodology of abstraction and conceptual construction, the conceptual framework, the method and the procedure of the objective evaluation of the user experience were defined. Finally, a double experimental evaluation process allowed to prove the feasibility of the proposed method.

Results: We obtained three main results. a) A conceptual Framework for User Experience Evaluation. Through this effort, we extracted a list of concepts of taxonomic order, from which we build the conceptual framework. We obtained the method from the concepts defined in the conceptual framework and organized it so that its information flow was in accordance with the guidelines set out in the structure of quality standards for the evaluation of software systems. b) A measurement procedure for User Experience Evaluation. We define a measurement procedure to find the relationship of the electrodermal activity signal with the psychological event cognition. c) Evaluation of the conceptual framework, the objective method and the measurement procedure. The evaluation entails: (a) qualitative results in the validation of the conceptual framework, (b) statistical analysis for the understanding of the data obtained from the recording of electrodermal activity signals, and (c) machine learning results with the generation of a classification model that relates in an acceptable way the data of electrodermal activity with the aspect 'Performance' of cognition.

Conclusions: The objective method for evaluating the user experience provides a logical, comprehensive and practical representation for the application of any measurement procedure in a given evaluation context of the user experience. This representation allows the implementation of non-intrusive physiological sensors to obtain knowledge about the relationships between the physiological responses of the user and any psychological aspect that the user experience evaluator wishes to know. The classification model generated from the EDA signals and the cognitive performance of the user provides an excellent basis for future work for the recognition of response patterns in users, particularly in children with cognitive disorders.

Keywords: User Experience, Evaluation, Conceptual Framework, Method, Procedure, Physiological Signals, Psychological Events, Electrodermal Activity, Cognition.

Resumen estructurado

Antecedentes: La experiencia de usuario pretende determinar, en un sentido amplio, el estado de las interacciones de un usuario con un producto, sistema o servicio. Particularmente para los servicios que el área de las Tecnologías de la Información y la Comunicación brinda al sector de la salud, la evaluación de la experiencia de usuario se convierte en un hito importante en el diseño e implementación de este tipo de servicios, con el fin de mantener la motivación del usuario al utilizarlos. Sin embargo, la evaluación de la experiencia puede ser intrusiva para el usuario y los resultados son sesgados por la subjetividad del usuario y el evaluador. Métodos de evaluación a partir de la psicofisiología dan cuenta de las relaciones que pueden existir entre lo que el usuario siente, piensa y hace, y lo que se puede medir en su cuerpo de manera no intrusiva, con el fin de lograr resultados objetivos en las evaluaciones de la experiencia de usuario.

Objetivos: El objetivo principal de esta tesis es la definición de un método para la evaluación objetiva de la experiencia de usuario para niños con trastornos específicos del aprendizaje. Para definir el método es importante referenciarlo en un marco conceptual que identifique las medidas psicofisiológicas utilizadas en la evaluación objetiva de aplicaciones en el ámbito de la eSalud. La instanciación del método definido lleva a la obtención del procedimiento de medición para definir la relación entre una señal fisiológica y un evento psicológico. El método es evaluado dentro del escenario de HapHop-Fisio para determinar la experiencia de usuario.

Métodos: Para alcanzar este objetivo, esta tesis se dividió en tres fases: en primer lugar, el levantamiento del estado del arte se realizó a partir de las metodologías del mapeo sistemático y la revisión sistemática, con el fin de obtener la base conceptual de este trabajo de investigación. En segundo lugar, por medio de una metodología de abstracción y construcción conceptual, se definió el marco conceptual, el método y el procedimiento de evaluación objetiva de la experiencia de usuario. Finalmente, un doble proceso de evaluación experimental permitió probar la factibilidad del método propuesto.

Resultados: Se obtuvieron tres resultados principales. i) Un marco conceptual para la evaluación de la experiencia de usuario. A través de este trabajo, se extrajo una lista de conceptos de orden taxonómico, a partir de los cuales se construyó el marco conceptual. Se obtuvo el método a partir de los conceptos definidos en el marco conceptual y fue organizado para que su flujo de información estuviera acorde a los lineamientos planteados en la estructura de los estándares de calidad para la evaluación de sistemas software. ii) Un procedimiento de medición para la evaluación de la experiencia de usuario. El procedimiento de medición se definió para encontrar la relación de la actividad electrodermal con el evento psicológico cognición. iii) Evaluación del marco conceptual, del método objetivo y del procedimiento de medición. La evaluación implica resultados: (a) cualitativos en la validación del marco conceptual, (b) de análisis estadístico para la comprensión de los datos obtenidos de la grabación de señales de la actividad electrodermal, y (c) de aprendizaje de máquina con la generación de un modelo de clasificación que relaciona de manera aceptable los datos de actividad electrodermal con el aspecto ‘Desempeño’ de la cognición.

Conclusiones: El método objetivo para la evaluación de la experiencia de usuario provee una representación lógica, completa y práctica para la aplicación de cualquier procedimiento de medición en un contexto de evaluación dado de la experiencia de usuario. Esta representación permite la implementación de sensores fisiológicos no intrusivos para obtener conocimiento acerca de las relaciones entre las respuestas fisiológicas del usuario y cualquier aspecto de la psicología que el evaluador de la experiencia de usuario desee conocer. El modelo de clasificación generado a partir de las señales EDA y el desempeño cognitivo del usuario provee una excelente base de trabajo futuro para el reconocimiento de patrones de respuesta en los usuarios, particularmente, en niños con trastornos de tipo cognitivo.

Palabras Clave: Experiencia de Usuario, Evaluación, Marco Conceptual, Método, Procedimiento, Señal Fisiológica, Evento Psicológico, Actividad Electrodermal, Cognición.

Content

Figure list

Table list

Chapter 1

Introduction

1.1. Problem Statement

In the last decade, the creation of Information and Communication Technology (ICT) services and/or products took a turn from the support of basic human tasks to the integration of the service into the user's lifestyle. Therefore, aspects non-inherent to the technological facet of the service such as satisfaction, enjoyment, and entertainment are part of the process of interaction of the user with the technology [1]. By placing this process in a specific context, together with the attribution of a sense of appropriation of the interaction by the user, the result is what is known as User Experience (UX). In consequence, the expectations, motivations, and feelings of the user when using an ICT service need an inquiry beyond traditional functionality and usability concerns [2].

A lot of the UX measurement tools can be found in the allAboutUX.org repository [3] and others in comprehensive reviews [4], [5]. This kind of tools depends on the collected data, the measures (being subjective or objective), and the way the data are collected [6]. The difference between the different measurement tools relies on those properties of the UX that can be reliably and repeatedly measured and those that cannot, like the things attached to an individual's background [7]. However, the distribution in the use of measurement tools is inclined towards subjective tools, i.e., methods that "measure" the user from his/her ability to externalize experiences or draw conclusions in light of observations made from an expert. Furthermore, those tools considered as objective, such as the ones using psychophysiological measures, tend to be intrusive for the user, which could prevent the user from having a natural behavior in the face of an interaction situation with ICT services [8].

The current scenario of this research comes from the HapHop-Physio project. HapHop-Physio was a technological transfer project developed at Universidad del Cauca in collaboration with the Fraunhofer Institute in Germany. HapHop-Physio is a serious application classified as an exer-learning game, and it supports the rehabilitation of children with intellectual and cognitive disabilities, focusing on memory and concentration therapies. While developing the game, it was particularly challenging to measure satisfaction in children, whether they would be able to play the game, do the therapies and have fun while using it. However, it was not possible to identify enough adapted tools to evaluate the satisfaction, and moreover, to evaluate the whole experience the user (i.e., the children) could have while using this technology [9].

This Master work focuses on children with some cognitive disorders grouped in the Specific Learning Disorders (SLD): known as alterations of the cognitive processes that interfere significantly in the performance of children in school. Experts gave complete diagnosis by identifying symptoms as dyslexia, dysgraphia and/or dyscalculia; there are also learning disorders not specified such as some behavioral disorders. Having appropriate educational support and therapies can help manage the SLD [10].

When evaluating UX of children with SLD, the objective methods for UX evaluation are relevant since they are not willing or able to respond questionnaires. However, as demonstrated in the state of the art section, there has not been found UX evaluation methods, especially design for rehabilitation or improvement therapies based on serious applications applied to children with SLD. Therefore, this Master thesis intends to respond the following research question:

How to support the objective evaluation of user experience of children with SLD using HapHop-Physio?

1.2. Motivation

The diagnosis and treatment of disorders associated with cognitive and/or behavioral development in school-age children are focused on enhancing basic cognitive skills for learning, such as memory, attention, spatial orientation, language, organization and planning (executive functions) and metalinguistic skills (such as spelling) [10].

Conventionally these therapies use as reference the activities evaluated in standardized batteries [11]–[13], whose resources are limited to pencil and paper along with the specialist guiding the child through the activities of the session. From the clinical experience of the specialists involved in the HapHop-Physio project (Nathalia Narváez, neuro-psycho-pedagogy specialist, and Luz Ángela Tobar, specialist in neuronal rehabilitation from Fisiocenter S.A.S.[1]), this situation has the same characteristics of the environment that children experience in school or at home, which could generate fatigue and lack of motivation in the child, affecting the progress and results of the session, and even generating the desertion of the therapies.

The HapHop-Physio system is an alternative in the application of memory and attention therapies, especially in the auditory component, in conjunction with the practice of reading and writing processes. Since the system has a mat as an input device, the interaction between the user and the game guarantees physical movement. Studies have shown that children with SLD can improve their learning through movement to maintain the level of alert needed to perform cognitive activities [14], [15]. From the point of view of the companies providing health services, HapHop-Physio represents an innovative product of clinical work and objective in the follow-up of the children. Having therapeutic control over the results of therapy sessions is an advantage for health professionals.

Having in mind the considerations regarding the target population and the type of intervention provided by HapHop-Physio, the evaluation of the UX is vital to corroborate through the objective monitoring of the child: if children like the therapeutic intervention,

[1] Healthcare Provider Institute registered to the Ministry of Health of the Colombian Government with the number 1900107122.

are interested in or not and, if they are motivated to do the therapies or not. Measurement and analysis of the UX allow giving an anticipated diagnosis of how the psychological situation of the children can affect their performance, as well as taking corrective measures for the improvement of the therapy. In this way, children will not deviate from the therapies and there will be an improvement in the cognitive and learning processes of the children.

1.3. Objectives

1.3.1. General Objective

Propose a method for the objective evaluation of user experience of children with SLD using HapHop-Physio.

1.3.2. Specific Objectives

1. Identify validated psychophysiological measures used to evaluate user experience in serious applications for health.
2. Propose a measurement procedure for the objective evaluation of user experience in children with SLD, which integrates one of the validated psychophysiological measures identified.
3. Evaluate the proposed method in the context of the HapHop-Physio application.

1.4. Contributions

The main contributions of this master’s thesis are:

- A Conceptual framework proposed to identify and associate the main concepts of UX, from the perspectives of the user and system. The obtained plane[2] provides a better overview of the UX phenomenon, clarifying the limits within it, i.e., what is measurable and evaluable from the user and the system – Chapter 3.
- An UX Evaluation Method built and defined from the relevant concepts within the UX field, taking as reference standards for the Measurement of Systems Quality. Such method allows us to plan and implement the evaluation of the UX, in particular, the momentary UX, in the most objective way possible – Chapter 4.
- A Measurement procedure defined from the instantiation of the method built to logically organize the objective measurement of the UX through obtaining and processing psychophysiological measures – Chapter 4.
- An evaluation of the feasibility of the objective method and measurement procedure in a scenario measuring changes in cognition of the children with SLD using the HapHop-Physio serious application (exer-learning game), supported by the Electrodermal Activity (EDA) data. – Chapter 5.
- Four papers published in indexed journals and international conferences.
    - “HapHop-Physio: a computer game to support cognitive therapies in children”. Journal paper. Published in *Psychology Research and Behavior Management* (Publindex A2). Paper relating the initial process of the HapHop-Physio application establishing the research gap that gives rise to this research work.
    - “User Experience Evaluations in Rehabilitation Video Games for Children: A Systematic Mapping of the Literature”. Conference paper. Published in *German Medical Data Sciences: Visions and Bridges* from the series *Studies in Health Technology and Informatics* (Publindex B). Paper with an overview of the state of the art in the UX field by following the systematic mapping methodology.
    - “User Experience Evaluation Methods in Games for Children with Cognitive Disabilities: A Systematic Review”. Conference paper. Presented at *International Young Researcher Summit on Quality of Experience in Emerging Multimedia Services QEEMS 2017* (Awarded as

[2] As a graphic representation [16]

Best Paper). A study complementing the state of the art following the systematic literature review methodology and the PRISMA statement for reporting secondary studies in health.
    - "Towards a Conceptual Framework for the Objective Evaluation of User Experience". Conference paper. Published in the *Lecture Notes in Computer Sciences* series (Publindex A2). Paper considering the creation of a new conceptual framework that identifies the differences between the UX evaluation perspectives and their measurable aspects.
    - "Using Electrodermal Activity to Recognize Cognition in Children with Specific Learning Disorders". Journal paper submitted to the IEEE Access Journal (JCR Journal). Paper relating the obtained data from the UX evaluation, validating the method built and the procedure implemented through the use of physiological data (EDA).
- Knowledge Transferred to Industry. The Master thesis strengthened the clinical services offered by Fisiocenter S.A.S, especially comprehensive rehabilitation programs for neurodevelopment in children. A neuro-rehabilitation station for children supported by ICT was deployed in the clinic facilities, and the differentiating value of this implementation lies precisely in providing accurate information on the psychological state in an objective way (through physiological data) of the child at the time of interaction with the HapHop-Physio application.

1.5. Contents of the Dissertation

This dissertation is divided into six chapters, which we describe as follows:

1.5.1 Chapter 1. Introduction

In this chapter, we start by describing the problem addressed, the proposed objectives to find a solution to the stated problem, the main results obtain during the development of this thesis and the contents of this dissertation document.

1.5.1. Chapter 2. State of the Art

In this chapter, we present a brief description of the most relevant concepts for this thesis. Moreover, we also analyze different research projects working on the UX evaluation in children interacting with a range of games through measurement methods to establish our research gaps.

1.5.2. Chapter 3. Objective Evaluation for User Experience

Throughout this chapter, we present a conceptual framework that identifies and associates the main UX concepts, from the user and system perspectives. We explain how we obtained the plane of concepts and how the built framework led to the definition of our objective UX evaluation method

1.5.3. Chapter 4. Objective Method for User Experience Evaluation

In this chapter, we present the objective method for UX evaluation where the physiological signals are the convergence point between the internal state of the user and the measurement of the cognition. Also, we introduce the process of instantiation of the method to obtain the measurement procedure. Likewise, we analyze different tools for recording physiological signals that allow evaluating the physical and psychological state of the child in a non-intrusive way. Finally, we expose the measurement processes that must be carried out from the defined procedure.

1.5.4. Chapter 5. Evaluation

Throughout this chapter, we present the evaluation of the conceptual framework, the method, the data collection, the analysis, and relevant information about the experiment that we conducted during the application of a pilot test with the HapHop-Physio system under the supervision of specialists in pediatric neuro-rehabilitation. Furthermore, we evaluate the data through statistical analysis and a machine learning approach to validate the proposed method.

1.5.5. Chapter 6. Conclusions and Future Work

Finally, from the implementation of our objective method and its validation process through the measurement procedure, we obtain the most important conclusions and propose different future works that could increase the impact of our UX evaluation method.

Chapter 2

State of the Art

In this chapter, we introduce the description of the relevant concepts for this thesis. In addition, we present the analysis and the research gaps found from studies about UX evaluation in children interacting with a range of games through different measurement methods.

2.1. Background

With the aim of offering a general background for this thesis, this section includes the explanation of three comprised concepts (build from other concepts) surrounding the research context.

2.1.1. Evaluation Methods from User Experience

The UX allows professionals who are around the creation of products and services to have a holistic vision to include non-functional elements within the system to provide experiences, stimulate sensations, and generate emotions such as joy, fun, and pleasure [17]. The evaluation of UX serves to support the selection of the best design, to ensure the proper development of the system (product or service), or to measure when the final product meets and satisfies the original objectives of the UX [2].

An evaluation method of UX is a method measuring UX aspects to get information about the fulfillment level of a certain aspect [5]. As defined in [6], we consider the evaluation method of UX as the actions to conduct the evaluation and discover how a person feels using a system, by means of collection techniques to gather data from the experience, and not only reduce the evaluation method to a simple tool or to perform a simple measurement of the UX.

There are studies [6], [4], [2], [5] that already identify and classify different evaluation methods found in the literature. Classifications are made according to the type of research/study, the development phase, the type of research conducted, the users and evaluators, the restriction of time, to the period of experience, to the data collected and even, to the place of evaluation.

2.1.2. Children with Specific Learning Disorder

The term "learning problems" arose in 1962 to refer to a delay or disorder in the development of speech, language, reading, writing or mathematical skills that result from brain dysfunction or emotional problems. Nowadays, emotional or social interaction problems have a relation but they are not the main cause of learning problems. The learning problems have an intrinsic nature due to their relation with a dysfunction of the central nervous system.

Specific learning disorders include dyslexia, dysgraphia, and dyscalculia, as well as problems in speech and language development, non-verbal learning problems, and difficulties in the development of motor skills.

Dyslexia is a specific learning disorder of neurobiological origin. Children suffering from this condition have difficulties in the accuracy and/or fluency in word recognition, as well as a lack of spelling and decoding skills. These difficulties result from a deficit in the phonological component of language.

The disorder in written expression within the group of learning disorders is known as dysgraphia. The diagnosis requires that writing skills are below expectations for chronological age, intellectual level, and the child's age/grade level relationship. Problems in writing interfere with academic performance and activities of daily living that require writing.

Finally, having dyscalculia causes difficulties in a variety of numerical tasks such as performing arithmetic operations, solving mathematical problems, and using numerical reasoning [10].

In general, there are different difficulties for children who have learning disorders as problems in the acquisition of skills in reading, writing and mathematics result in low academic performance. Another type of difficulty is the communication disorders characterized by speech or language impairments, including expressive language disorder, stuttering and other communication disorders where there is an alteration in the articulation, expression or understanding of language. Finally, generalized developmental disorders present alterations in multiple areas of development. It includes alterations of social interaction (in the game and integration with other children) alterations in communication (not expressing emotions) and the presence of stereotyped or repetitive behaviors and activities [18].

2.1.3. Serious Applications

Serious applications can be any type of digital systems used for different purposes than generating entertainment. Serious applications are also called "serious games" and usually refer to games used for training, advertising, simulation, education or health, which are designed to run on personal computers or video game consoles [19].

However, serious games are not merely the application of regular games technology for non-entertainment purposes to domains such as education and health; these games should be more attractive and motivating to benefit the development of learning skills and healthy habits in the users [20].

In the end, a serious game is the application of technology, the process and design of games to the solution of problems faced by people to organizations. Serious games promote knowledge transfer and game development techniques in traditionally-not-related areas to games [19].

These types of applications have a positive impact on the development of a series of different skills by players/users, like motor skills and spatial recognition [21], [22], educational and formational effects [23]–[25], and the improvement of social and emotional skills [26], [27].

2.2. Related Works

To find previous works reporting results concerning the evaluation of the UX in serious applications that support the rehabilitation processes of children with SLD, we followed two methodologies for a state of the art analysis. The first methodology is the "systematic mapping" where we found general but relevant information about the type of users of the serious applications in the health domain and the evaluation treatment carried out when obtaining UX [28]. The second methodology is the systematic review of the literature. According to the previous results, we carried out the deepening of the most relevant works that contained the greatest amount of elements that would approach a possible solution to the problem we found [8]. This section contains the results statement obtained from the application of both methodologies.

- **UX Evaluation**

When discussing the evaluation in UX within the field of Human-Computer Interaction (HCI), we recognize two important perspectives of this evaluation. First, we have the evaluation of the system to improve it, and second, we evaluate the user to improve the system. Both perspectives involve the user: the first one is about what the user says about the system, and the second one is about how the system impacts the user.

According to the studies found, not all of them use the existing UX evaluation methods to evaluate the designed/developed system where most of the studies used ad-hoc evaluation methods. In Figure 1, the studies evaluating the user or the system are detailed and if the evaluation method used was based on the UX methods or not.

The user evaluations that we identified in the studies were, among others, controlled trials, taking neuropsychological measures, obtaining psychophysiological data, and recording audios and videos. Besides, evaluations made on the system included case studies, comparative studies, and feasibility studies. Most of the UX evaluations carried out were subjective (82% of the studies). Most of the studies that conducted evaluations used subjective measurement tools.

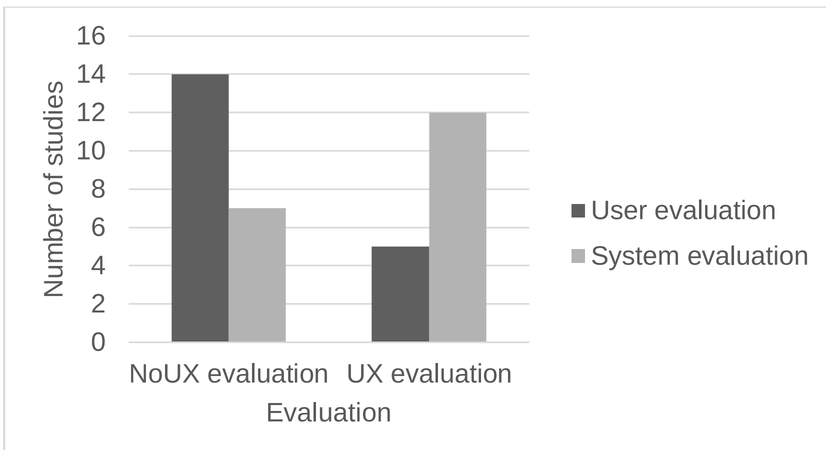

Figure 1. Evaluations performed in studies

The studies that conducted objective evaluations used multiple devices to take physiological measurements to obtain objective data, regardless of whether they were evaluating the UX or not. Some of the physiological measures mentioned in the studies include EEG brain waves, electro dermal activity (EDA) [29], stress levels, contraction of the facial muscles [30], movement and postural attitudes, galvanic skin response (GSR) [31], skin temperature, heart rate (HR), interbeat interval (IBI), heart rate variability (HRV), and respiratory rate (RR) [32].

- **Target Users**

Among the preliminary results that we obtained in the application phase of the systematic mapping methodology, we found that although the majority of the serious applications with rehabilitation purposes have children as target users, 13% of the studies directed their efforts towards the population of the third age [33]–[37], [22]. There were some studies that did not report the target audience [38]–[40], [25], however, we concluded it could be adolescents, adults or anyone who suffered a sudden accident.

We would like to notice that not all the studies were designed and/or developed for children with cognitive deficiencies. Almost half of the studies (41% of them) sought to improve cognitive skills for children in their stage of development [41], [23], [30], [42]. In addition to SLD [43], we found other disabilities in the studies: children with ADHD [44], [45], Down syndrome [46], obsessive-compulsive disorder [47], autism/Asperger [48], [49], and cognitive disabilities in general, such as delays in speech [50], and even children with HIV vulnerable to impairments in the cognitive area [51].

- **Serious Applications**

From the search performed, it was possible to identify seven types of applications, including video games and serious games, however, the type of learning games through exercise was not found (HapHop-Physio belongs to this classification). The category of Human-Computer Interaction systems (HCI) includes different types of systems: games, interactive games, web platforms, technological-solutions systems, robots, brain-computer interface systems, and haptic systems (see Figure 2).

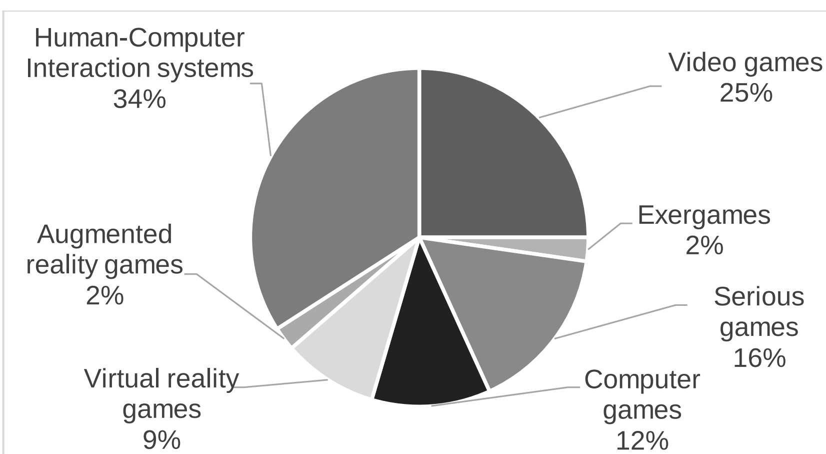

Figure 2. Serious applications types

When observing the serious applications (especially the games) we could recognize its wide use for children due to its power of entertainment. In the same way, elderly population finds a solution to degenerative diseases or loneliness [34], [36], [37].

- **UX Evaluation of Children with Cognitive Disabilities using Serious Applications**

Taking into account the individual findings, we carried out the systematic review of the literature to expand the information on the evaluation of UX in serious applications for children with cognitive disabilities, especially with SLD. We summarize most of the results obtained in this review in Table 1. We present in detail the measurement tools used to determine the UX (classification made from *www.allaboutux.org*) according to the system that authors of the studies developed and presented, together with the elements considered to determine the results of the evaluation and their respective subjective or objective measures.

Table 1. Studies performing evaluation methods from UX

Ref.	UX measurement tools	Subjective measures	Objective measures	Evaluation aspects	Application
[52]	Controlled observation, extended usability testing, and game experience questionnaire.	✓		Attention span, visual perception, visual memory, comprehension, spatial orientation, visual discrimination, social environment, learning level, degree of hearing loss, motivation, effectiveness, efficiency, satisfaction, emotions and learning.	Serious game for children with hearing impairment with learning purposes.
[53]	UX curve and controlled observation.	✓		Linear positive, linear negative and non-linear responses.	Games for healthy children to evaluate UX.
[54]	Controlled observation and	✓	✓	Reasoning, attention, memory,	Serious application

	psychophysiological measurements.			spatial memory, executive function, cognition function, emotions, episodic memory, visual memory, orientation, spatial-frequency, and visual world.	for ADHD children and children with SLD having a learning purpose.
[55]	Controlled observation, TUMCAT, Wizard of Oz.	✓	✓	Usability, child preferences, gestures, fun, and social interaction.	Games for healthy children with learning purpose.
[56]	Playability heuristics.	✓		Do not mention evaluation aspects.	Videogame with neurofeedback for ADHD children with rehabilitation purpose.
[57]	User experience questionnaire adapted and controlled observation.	✓		Attractiveness, efficiency, perspicuity, dependability, stimulation and novelty.	Mobile application for healthy children to evaluate UX.
[58]	Psychophysiological measurements.		✓	Heart rate, pulse.	Active videogame for healthy children to evaluate UX.
[48]	Controlled observation	✓		Engagement, emotion, motivation, independence, communication, and creativity.	Kinect game for autistic children to evaluate UX.
[59]	Fun toolkit and This-or-that.	✓		Fun and ease of play.	PC and tablet games for healthy children to evaluate UX.
[60]	Controlled observation and psychophysiological measurements.	✓	✓	Interest, slight confusion, joy, feeling, of control, fear, strong confusion, anger,	Educational game for healthy people.

				boredom, and disregard.	

38.4% of the studies report the use of questionnaires, direct observation and self-report to evaluate the UX, being these considered as subjective tools. In addition, 30.8% of the studies used objective methods such as recording and analyzing the log of the system, or taking physiological measures such as the EEG. Finally, 30.8% of the studies use video recording, an objective tool that does not guarantee the total objectivity in the results because the analysis of its content is performed by an expert to reduce the risk of bias. Regarding the combination of evaluation methods of the studies reviewed, 40% showed a combination of subjective and objective methods. Only 20% used purely objective methods; the remaining 40% used only subjective methods.

From the studies reviewed, we highlight the study carried out in [54]: the KAPEAN tool helps to collect subjective and objective information from children suffering from ADHD and SLD. They collect data while children play with mini-games to improve their attention levels. The researchers demonstrate the attention level improvement of the children. However, the tool and mini-games were not suitable for all children. Authors did not report whether the children enjoy or have fun in the experimental sessions, and the evaluation tools used by the authors were intrusive for children as they were using an EEG headset. Therefore, from the UX perspective, they could not achieve the UX objective completely.

Finally, regarding the general purpose of the UX evaluation, the main objectives are related to (i) design solutions for a population, (ii) evaluate a solution, (iii) assess the user, (iv) improve a solution, and (v) improve the conditions of the user. It is important for the UX researcher to identify the main purpose of the evaluation and take a course of action leading to create standardized processes and methods.

2.2.1. Shortcomings

From the results obtained by reviewing the literature, we identified the following research shortcomings:

- We found that objective evaluation methods [59] are still lacking to evaluate UX in serious applications with rehabilitation purposes for children with SLD.
- There is an overlapping or undefined limit between the contributions made in the HCI and the health/health informatics domains. Some studies evaluate the user to show the functionality of the game or therapeutic impact of the technology [23], [48], [61], [51], while others evaluate the game to ensure user's acceptance [62], [57], [50].
- Most serious applications try to make user-centered designs or user evaluations. However, we do not find any application that covers the entire development cycle, from the design stage to the evaluation to take measures to improve the status of the application [33], [36], [22], [39], [25], [47], [63], [64].
- Most studies choose to evaluate the solution they developed and stay with the positive results. We did not find cases of implementation of systems capable of adapting to constantly improve the system [30] and, therefore, maintain the motivation of the user and generate adhesion to the system.
- The studies that reported the implementation of solutions for children with learning difficulties looked for the improvement in the educational level of the children, without considering the aspects of rehabilitation that could be carried out to improve the health status of the children [54], [56].

The addressed shortcoming in this Master’s thesis was the one related to the lack of objective methods for children’s user experience evaluation of serious applications for health.

Summary

In this chapter, we described the most relevant concepts for the foundation and understanding of this thesis:

- Evaluation methods from UX, where one or more measures can help to define if the final product meets and satisfies the original objectives of the UX and to get information about how a person feels using a system.

- Children with SLD are children having problems in the acquisition of reading, writing and, mathematics skills resulting in low academic performance. It is manageable with memory and attention therapies.
- Serious applications, application of technology, the process, and design of the solution of problems, in domains such as education and health, different from generating entertainment.

Moreover, through the analysis of the results from the systematic mapping and the systematic literature review, we identified an overview on the applications to support cognitive therapies for children and how to evaluate them using the UX methods. Likewise, we deepen in the identification of the evaluation elements and aspects of the UX methods used in the found studies. Such characteristics gave a perspective about the most used methods for assessing children in a subjective way. Particularly, we were interested in identifying subjective and objective methods in the UX evaluation. By last, we listed the gaps we identified in the reviews; our final goal is to determine an objective method for evaluating UX in HapHop-Physio application for children with SLD.

We obtained feedback from academic pairs by reporting the mapping and the review in two different conference papers. These presentations allowed us to refine the results, procedures and the methodologies followed.

To have a comprehensive foundation of concepts to obtain the first contribution from this Master's thesis, we took several studies from this state of the art; we describe the conceptual framework to evaluate the UX in the next chapter.

Chapter 3

Objective Evaluation of User Experience

For this chapter, we considered that the concept of UX involves many different aspects and perspectives. Then, we explain the relevant concepts of the UX to continue with those that belong to the evaluation of it. With these concepts in mind, we proposed a conceptual framework based on a methodology to identify the differences between the UX evaluation perspectives and their measurable aspects.

3.1. User Experience Concepts

The term UX appeared more than a decade ago becoming, after a few years, a phenomenon, a field of study and a practice [65], a core for experts and researchers in the field of HCI, and even penetrating other areas of knowledge such as psychology [47], medicine [66], learning [43], and advertisement [67]. At first, researchers and practitioners dedicated a few years to obtain a common definition of what is UX, what does it imply, how is it done, and what is its scope [68]–[70].

For defining the scope of this emerging area, researchers and practitioners took into account different standpoints: (a) usability before UX, (b) UX beyond completing tasks,

(c) UX as an emotional aspect, and (d) designing and evaluating for UX. We explain each one of them.

- **Usability before UX**

The traditional usability framework focused primarily on user's ability to understand (user cognition) and to use (user performance) an 'artifact' in HCI [71]. Usability remained as a necessary condition in the context of interactive products and software, however, "it is not sufficient to make a user happy". In accordance, most researchers argue that UX emerged as an extension of usability to accommodate the fuzzy quality attributes of experience such as enjoyment, pleasure or fun, whereas others argue that UX exceeded usability by including it [68].

- **UX beyond completing tasks**

The pragmatic/hedonic model from Hassenzahl [17] gave the UX two dimensions for measuring it, differentiating from a focus on the product (pragmatic quality) and a focus on the self (hedonic quality), i.e., on the subjective side of the product [72]. This model links product attributes with human needs (i.e., personal growth, self-expression, and self-maintenance) and values (i.e., increasing knowledge, skills, and memories) [73].

- **UX as an emotional aspect**

There is a common understanding of the holistic nature of UX as it makes emphasis on the emotion, motivation, and action, in a given physical and social context. In addition, UX is subjective; it focuses on the "felt experiences" rather than the product attributes [74]. In UX, there are two basic ways in dealing with emotions: one emphasizes the importance of emotions as consequence of product usage; the other line focuses on their importance as previous circumstances of product usage and evaluative judgment [73]. Thüring and Mahlke's framework [75] explicitly defines emotional reactions as an integral component of the user experience and not as a consequence. On the other hand, Mandryk's approach explored physiological data as a direct indication of UX through mathematically modeling emotion [76]. Both models measure physiological reactions; these measures are valued for being unobtrusive and therefore able to monitor, in a constant way, indices of emotion, instead of asking the users to stop their experience, reflect on their emotion, and then disclose it [77].

- **Designing and evaluating UX**

Both design and evaluation methods have interest in finding ways to evaluate UX of current concept ideas, design details, prototypes, or final products [2]. The primary effort of evaluation methods is to support and help in selecting the best design so development is on the right path, or to measure whether the final product meets and comply with the original UX targets [78]. Kort's framework studies the sense-making process, the UX aspects, and their relationships with design elements intended to create specific experiences [1]. However, this is not an operational framework.

Besides these standpoints, the ISO 9241-210 [79] standardized the concept of UX in 2010. With the arrival of the ISO standard and previous work on handling the standard since 2009 (e.g., [80], [81]), researchers and practitioners from the HCI community were divided. There are those who accepted the norm, stuck to it and still do it, and those who, based on their experience in academia and industry, built their own definition and work(ed) around it, always keeping in mind that the concept of UX is complex [65]. New researchers in UX area see in the ISO definition a safer way to start walking the long road built to date of what UX is and how we should do it.

The ISO 9241-210 standard, which was last reviewed and confirmed in 2015, defines UX as a "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service". This standard also clarifies that: first, UX includes the emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments from users and they can occur before, during and after use. Second, UX is a consequence of three factors:

- The interactive system characterized by the brand image, presentation, functionality, system performance, interactive behavior and assistive capabilities.
- The prior experience, attitudes, skills, and personality as causes for the user's internal and physical state.
- The context of use.

Third, usability criteria, interpreted from the perspective of the users' personal goals, can use perceptual and emotional aspects from UX to assess it [79]. As we have shown, UX involves many aspects and different perspectives, making it difficult to evaluate the whole set of what UX represents.

3.2. User Experience Evaluation Concepts

To the best of our knowledge, there is no clear definition of the UX evaluation concerning the identification of the different evaluative aspects according to the perspectives, and even taking into account the final needs of UX evaluation according to a given context of use. Thus, as a first step towards a clearer definition, we identified the differences between the evaluation perspectives and its measurable aspects.

We started by analyzing the UX evaluation domain from two perspectives: first, the system (product or service) and second, the person or user. From the system's perspective, there are different approaches to evaluate the UX:

- Several authors declare to perform UX evaluations when applying usability tests during verification processes of software quality [28].
- Miki's evaluation framework of usability and UX [82] tries to separate objective measures represented in effectiveness and efficiency from subjective measures represented in perceptions and satisfaction.
- The author in [80] discusses, analyzes and conceptualizes in three different ways the relationship between usability and UX: (i) UX as an elaboration of the satisfaction component from usability, (ii) distinct from usability which emphasizes in the user performance, and (iii) UX being an umbrella term to usability and other considered aspects in the HCI.
- In the software engineering standards, usability reduces the UX to a simple aspect of it. Consequently, the pleasure measure found in ISO/IEC 25022 [83] is equivalent to UX and the recommended methods to find out user's pleasure are questionnaires.

From the user's perspective, the main categories of the measurement methods to evaluate UX are:

- Observational methods, where an expert (UX researchers, psychologists, sociologists, anthropologists) estimates the user's reactions.
- Psychometric scales, in which the user evaluates himself/herself in a qualitative way. UX practitioners consolidate this information through scales to obtain quantitative data.
- Psychophysiological measures, measuring the body's responses to determine user reactions.

In the first two categories, we consider the methods as subjective, while in the third, the methods are objective [84].

3.2.1. System’s Evaluation Concepts

We use computer systems in an extensive variety of application areas, and their intended and correct operation is crucial for almost every human activity. Perform a comprehensive specification and evaluation of systems and software product quality is a key factor in ensuring adequate and high quality. We can achieve this task by defining appropriate quality requirements while taking account of the future use of the systems and software products [85].

The general goal of the Systems and Software Quality Requirements and Evaluation (SQuaRE) series of standards is to cover three complementary processes in systems and software: requirements specification, measurement, and evaluation, by assisting those developing and acquiring system and software products with high-quality requirements. Within the SQuaRE series, there is the Quality Measurement Division (ISO/IEC 2502n) that provides information and guidance about how to measure the characteristics and sub-characteristics of a quality model.

The associated standards and technical reports within the Quality Measurement Division describe measures of quality throughout the product lifecycle. Internal measures characterize quality based upon static representations of the software, external measures characterize quality based upon the behavior of the computer-based system including the software, and quality-in-use measures characterize quality based upon the effects of using the software in a specific context of use [86].

Bearing in mind the UX concept, we are interested in the quality-in-use requirement, defined as the “degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, and satisfaction and freedom from risk in specific contexts of use”. These are outcomes of interaction between a user and a system [83].

3.2.2. User’s Evaluation Concepts

Since the analysis of the ISO UX definition provide us a good basis for a better understanding of UX, we take a deeper look into some of its main concepts, beginning with the “person’s perceptions and responses" from the perspective of psychophysiology. As entries of a psychophysiological process, we have the perceptions occurring in the nervous system [87], and the responses being the outcomes of this kind of process.

More precisely, psychophysiology is a research field that tries to understand the underlying mental reactions/functions (psychological responses) connected to the body signals (physiological responses) [88]. The Central Nervous System (CNS) and the Peripheral Nervous System (PNS) control the physiological responses: the CNS manages all the information received from the body, whereas the PNS includes all nerve cells external to the CNS. The PNS transmits most of the physical sensations; therefore, on the skin is easier to measure these reactions [89]. On the other side, responses related to emotions and mental workload of the users are the psychological ones, in its most basic definition [76].

We consider these concepts as objects or entities to build the conceptual framework. The concepts are related to each other and we represented these relationships by concept taxonomies that define their hierarchy. In Annex A, we introduce the formal definition of each concept and the taxonomy built around each set of concepts.

3.3. Conceptual Framework for User Experience Evaluation

In this section, we briefly explain the construction of the conceptual framework based on a methodology. The section has two parts: first, we detail the methodology followed to analyze and synthesize the discovered information into the conceptual framework; second, we define the main concepts in our framework by analyzing the domain of UX evaluation, taking as reference the ISO 9241-210 standard and relating and introducing other concepts derived from psychophysiology.

3.3.1. Methodology for Building Conceptual Frameworks

According to [16], a conceptual framework is a network of interlinked concepts providing a comprehensive understanding of a phenomenon, and it is not a simple collection of concepts; each concept plays a representative role. This methodology presents eight key phases:

1. Mapping the data sources
   We mapped different sources of data according to the time division created by the ISO appearance:
   - Before ISO: data from workshops and special groups of interest from HCI conferences such as ACM CHI Conference on Human Factors in Computing Systems.
   - After ISO: the standard itself and papers with direct references to it.
   - Scientific databases like Scopus and ACM digital library for reviewing primary studies.
   - Papers on secondary studies [90], [8], [5], [28], [84].
   - The SQuaRE series of standards [85].
2. Categorizing the selected data
   We took into account four aspects to categorize the information found in the data sources:
   - UX design

- UX evaluation of the user
- UX evaluation of the system (product or service)
- Quality evaluation of the system

3. Identifying concepts
   We identified the concepts in both explicit and implicit ways, naming those concepts that were “discovered”.
4. Categorizing concepts
   In this phase, we organized the main attributes, characteristics, assumptions, and roles from the concepts according to common features.
5. Integrating concepts
   The aim of this phase was to integrate and group together concepts that have similarities through association of concepts or relationships.
6. Synthesis and resynthesis
   The theorization process of the grouped concepts is iterative and includes repetitive synthesis. From this phase we were able to start recognizing a general framework that made sense.

Regarding the seventh phase (validating the conceptual framework), we performed a validation through a research paper discussed in the 20th International Conference on Human-Computer Interaction. The discussion with practitioners/researchers helped us to receive feedback. We present the validation of the framework in Chapter 5. Therefore, we execute the eighth phase (rethinking the conceptual framework), knowing that UX as a phenomenon is a dynamic area within the HCI field. We also consider the evaluation proposed in the third specific objective as a way of validating the built framework.

3.3.2. Conceptual Framework

As part of the initial thinking of the proposed conceptual framework, the concepts and their associations took part in a plane facilitating the preliminary identification of the basics of the evaluation process in UX (Figure 3). Three sources acquired high relevance in the building of the framework: the ISO 9241-210 [79], the SQuaRE standards for Quality Measure Elements [91] and Measurement of quality in use [83]. Additionally, the research field of psychophysiology was relevant.

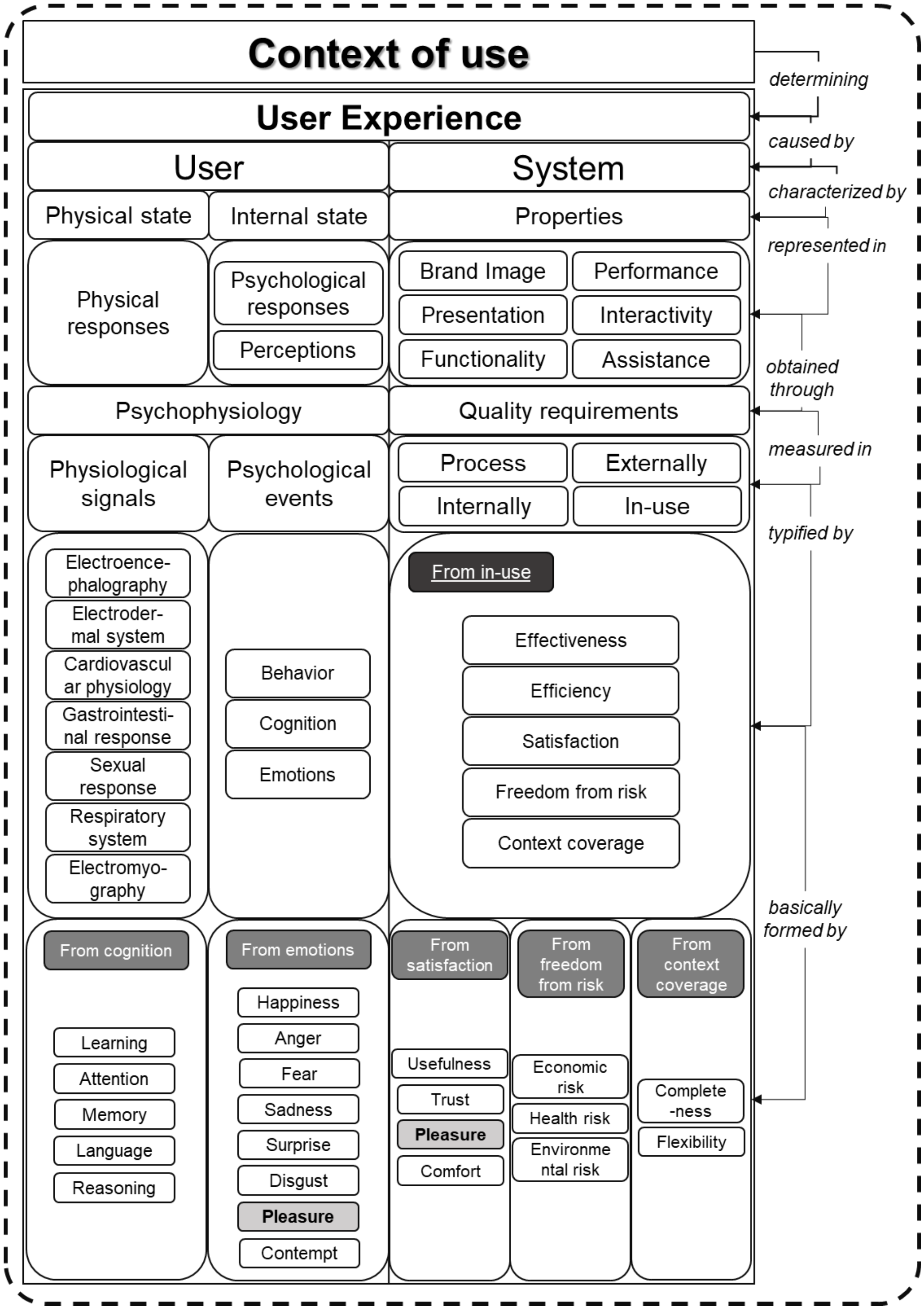

Figure 3. Conceptual framework for UX evaluation

The clarifying notes described within the definition of UX from the standard mentioned in section 3.1., helped in the consolidation of the framework, especially in the designation of the causal relationships between the key concepts. Thus, in the first part of the proposed framework, we declared that user and system are the cause of the UX, particularly the user's states and the properties of the system.

A third cause of the UX comes from the context of use. We omitted it to describe this perspective within the conceptual framework, given that we assume the UX evaluation as being made or performed on tangible ‘assets’, that is, the user and the system, and although the context of use must have the same importance as the other perspectives, we consider it as something intangible and therefore, not measurable over the ‘real world’. It is also important to consider that we see the context of use as the interaction (not tangible) therefore as determinant defining the UX. Physical and psychological responses and perceptions represent the user’s states (internal and physical) as well as the six properties assigned to the system.

From the user’s perspective, psychophysiology collects these states. As psychophysiology is the “scientific study of social, psychological, and behavioral phenomena as related to and revealed through physiological principles and events in functional organisms” [88], we concluded it gathers measures of physiological signals and psychological events separately, in order to obtain its counterpart. Nowadays, psychophysiology emphasizes their research in the map of the relationships between physiological responses and psychological events; a convenient way to understand the potential of these relations is to consider these two groups as representing independent domains within five general relations:

- One-to-one relation: a psychological response is associated with one and only one physiological response, and vice versa.
- One-to-many relation: one psychological response is associated with many physiological responses.
- Many-to-one relation: two or more psychological responses are associated with the same physiological response.
- Many-to-many relation: many psychological responses are associated with two or more physiological responses.

- Null relation: there is no association between the psychological response and the physiological response.

The most common case when performing physiological evaluation is where one body response could be associated with two or more mental effects or processes (the many-to-one relationship) [89].

According to Bernhaupt, from an HCI position, the overall goal of UX is to understand the role of affect as an antecedent, a consequence and a mediator of technology; hence, the UX concept focuses rather on emotional outcomes [92]. Bearing this in mind, from psychological events such as behaviors, cognition and emotions, we approached emotions from their own theory: we represented emotions in our framework inside the set of the 'basic emotions' defined by Izard [93] and Ekman [94]: surprise, sadness, anger, disgust, fear, contempt, pleasure and happiness.

However, given the final goal of the interaction between HapHop-Physio and the user, cognition becomes a better approximation from the point of view of objectivity. Cognition is defined as the ability to process information through perception [95] and include different cognitive processes or skills such as memory and attention [96].

It is important to highlight that we can also find some of these concepts used in the framework in the usability definition: “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” [79], since this concept comes from the HCI.

Finally, in our conceptual framework, we reached a common point between both perspectives; this common point is the emotion and quality sub-characteristic “pleasure”. Looking at it from the user's perspective, it is just another emotion included in a large set of variables we can measure from the user and to take into account in the UX evaluation. On the contrary, digging a little deeper into the sub-characteristic from system’s perspective, pleasure is the one variable defining all UX.

- **Implications of the Conceptual Framework**

There are considerations we need to highlight in relation to the construction of the conceptual framework:

1. The proposed framework differentiates between the elements involved in an interaction: user and system. Additionally, it considers the context of use. Within HCI community, some researchers and practitioners tend to overlap concepts, which, although they are difficult to separate due to their dependent nature, it is important to recognize their individualities and strengths in their singularity.
2. This framework relates the different forms of UX evaluation from the user's perspective. We take the process of expansion of the aspects to be evaluated in the UX to its "most basic" form (based on the selected concepts), allowing us to think about a whole lot of exploratory combinations to obtain the desired measurement. We can think about the observation of behavior, emotions or analysis of cognition through psychological methods.
3. It relates the standards to determine the quality of the system from the user's point of view because we base its conception on cascading concepts with general causality relationships. The abstraction levels of the framework help in the easy connection of concepts of the user with its counterpart in the system to form the sense of the framework.
4. The framework provides clarity regarding the procedure followed to evaluate the UX in a given context. That is the reason why we do not consider the evaluation of UX from the perspective of the context of use. Everything changes when defining that variable and we complement it with the recognition of the specific needs of each researcher and its project.

Summary

We proposed a conceptual framework in which we see in a more clearly and precisely way, the differences raised in the phenomenon of the UX evaluation seen from the two considered perspectives: user and system, under consideration of the context.

Likewise, the framework organized the most relevant concepts within the UX field, taking as a reference several standards and relating other areas of knowledge.

We followed a qualitative method for building conceptual frameworks. The conceptual framework has several implications for the UX community. It (i) helps in differentiating between the elements involved in an interaction: the user and the system. (ii) Relates the different forms of UX evaluation from the user's perspective. (iii) Relates the standards to determine the quality of the system from the user's point of view. (iv) Provides clarity regarding the procedure followed to evaluate the UX in a given context.

This approach attends as both conceptual and experimental basis for a measurement procedure that allows to objectively obtaining the UX in children with specific learning disorders, by measuring their physiological responses at the use time of a video game for cognitive rehabilitation.

From the proposed framework, we obtained an objective method to evaluate the UX in the context of the HapHop-Physio application. In the next chapter, we present this method along with the procedure followed to obtain the data that support the validation.

Chapter 4

Objective Method for User Experience Evaluation

In this chapter we address one of the contributions of this research: the objective method. We explain the function of the conceptual framework for user experience evaluation in the construction of the method [97]. Next, we instantiated the method towards what we have called the measurement procedure, a way to make the proposed method more intelligible. Finally, we relate the first steps in the course of action to follow for the complete application of the procedure.

4.1. Measurement Method

As mention in Chapter 1, HapHop-Physio is a project developed to support rehabilitation therapies for children with SLD. Through a video game of memory and concentration, we encourage children to complete the therapies while having fun. Evaluation of the UX is particularly important in ICT solutions for health such as this one since users/patients must maintain the motivation to continue using the technology and increase the adherence to their treatments or interventions.

Given that our users are children in cognitive rehabilitation processes, the UX evaluation would become a key factor in understanding the results of the child's therapy sessions. However, the measurement of UX aspects is a difficult and exhausting task for test participants. With methods such as psychometric scales (like Fun toolkit or This-or-that [59]), children are easily distracted, not impartial in the evaluation, and their reflection processes are not mature enough, which would prevent the externalization of their feelings in a rational and accurate way. Besides, regarding the use of observational methods, children tend to reject what is intrusive and might inhibit their natural performance.

The conceptual framework (Chapter 3) helped us in the definition of this objective method for the UX evaluation because it allows us to observe the dynamics of the UX phenomenon in a clearer and simpler way, building a method that considers a real application scenario.

Therefore, and for our case, we need an evaluation method that takes into account children as end users, that we need measurements and responses as close as possible to objectivity (without risk of bias), and that allows children to behave naturally in an environment closest to reality. In an extensive search of repositories and compilations in the literature of evaluation methods in UX, we found that psychophysiological measures are the closest to cover our needs. Nevertheless, most studies using these types of evaluation methods only consider adult populations, and the measurements are intrusive and sometimes annoying for users [28].

Searching for an objective evaluation method of UX, we found a guide that interrelates concepts of quality measurement in a simple and understandable diagram in the ISO/IEC 25022 standard [83]. By equating the concept of ‘quality’ with the appropriate concepts and relationships referenced in the conceptual framework we have a starting point for creating our method (Figure 4), considering the lack of any other method suitable for us.

Our method is simplistic: physiological signals become the convergence point between the internal state of the user and the measurement of the skills from cognition, such as memory or attention. The proposed method has a section division, according to the aim we want to achieve, and an element division, according to the representation of each element within the method.

Regarding the sections division, as we already established in the earliest stage of this work, we needed to find a way to measure UX, and in the conceptual framework, we were able to define cognitive abilities (a categorization of cognition) as a subset of the evaluative components for UX. Therefore, the first section of the method represents what we want to know, and the second section represents the actual method to discover it. Moreover, there are two representations when describing the method elements: the elements in rectangles as objects or entities, and the elements in ovals as processes or operations. This differentiation allows delimiting the results that we can obtain with the application of the method.

In this objective method (Figure 4), the internal state of the user includes those measurable experience properties associated with the psychological events. The physical responses measure these properties. These responses are from the processes that occur in the CNS and constitute a reply or a reaction due to the application of a stimulus [98]. A physiological signal measures the result of a physical response.

The relationships defined by the psychophysiology precise the elements within the section of the method that we want to know. Psychophysiology emphasizes their research in the map of the relationships between signals and cognition. A convenient way to understand the potential of these relations is to consider a many-to-one relationship, where one body signal is associated with the cognitive activities performed by the user (in our case) [88, p. 9]. The result of discovering these relations is obtaining the psychological events. In this way, psychological events become quantifications of the cognitive skills like memory and attention. It is important to highlight that our method also allows measuring the UX through other psychological events.

This method becomes the conceptual basis for the exploration and subsequent implementation of a measurement procedure and the experimental basis that allows objectively obtaining the UX in children with SLD, by measuring their physiological responses at the usage time of HapHop-Physio.

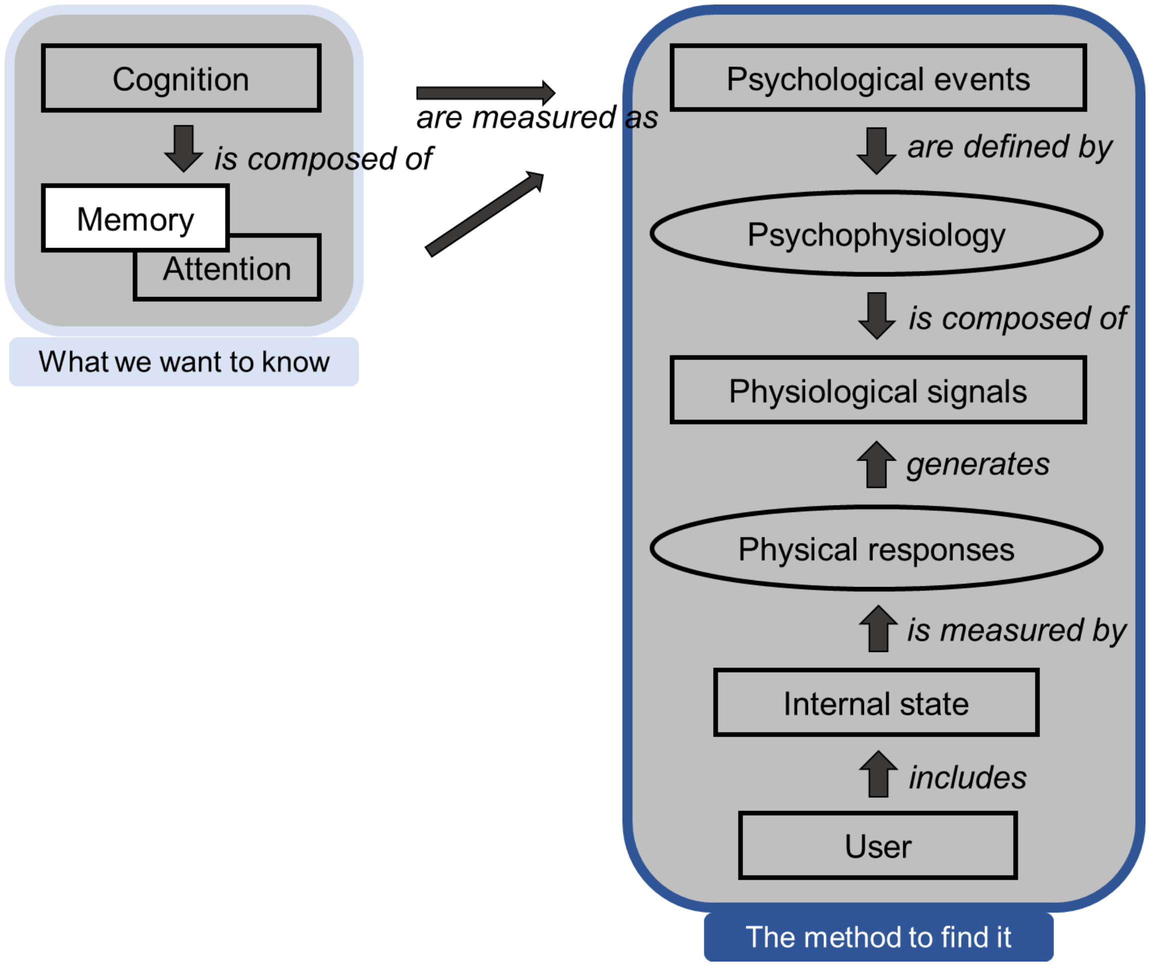

Figure 4. Objective method for the UX evaluation. Adapted and modified from ISO/IEC 25022.

In the next section, we present the instantiation of the objective method within our usage scenario, by mapping one physiological signal to obtain an objective approximation of the user's internal state.

4.2. Measurement Procedure

The measurement procedure is a logical organization of operations, applied in a specific way, used to perform measurements according to a measurement method [91, p. 3]. We consider the procedure as the instantiation of the objective method that particularly defines each one of the entities and operations from the method. As we observed in the method, the entity 'physiological signals' is the actual convergence point of all its elements. Thus, the first task towards the definition and implementation of the procedure lies in the selection of one of the signals considered in the conceptual framework.

4.2.1. Review of the Physiological Signals

In this subsection, we describe the physiological signals considered in our framework. For a better understanding, it is important to take notice of the general picture: what we call as 'physiological' refers to the functions in the body [99]. The nervous system controls these functions, through its divisions into the central nervous system (CNS) and the peripheral nervous system (PNS).

For reasons of a complete exploration of most of the options considered as evaluation components in the conceptual framework, we included signals of the gastrointestinal response and sexual response; however, we do not take into account these signals for the selection of the physiological signal due to the target users the ease of access to the measurement sensor.

- **Central Nervous System**

The CNS manages all the information received from the body and coordinates its activity accordingly. The bones of the skull and spine protect the CNS, which also hinders access from outside the body [100]. Here we show the most popular physiological signal for measuring electrical activity in the CNS.

a. Electroencephalography – EEG

EEG uses electromagnetic measurement to acquire the electrical activity within our brain. It is less invasive and easier to use compared to other CNS response analyzes systems, such as functional magnetic resonance imaging (fMRI), which analyzes brain function based on the hemodynamic response, or positron emission tomography (PET), which uses metabolic response to show the brain function [101].

EEG measures mild electrical activity, such as signals generated by neuronal activity in the brain [100]. EEG measures the amplitude of this activity in dimensions of a few microvolts (μV), generally between 10 and 100 μV.

EEG devices used for research, process the brain waves in different frequency bands, such as alpha (8-13 Hz), beta (13-30 Hz), theta (4-8 Hz), delta (1-4 Hz), and sometimes gamma (30-50 Hz) [102]. Alpha activity indicates a drowsy or relaxed state and lack of active cognitive processes; beta activity replaces alpha rhythms during cognitive activity and is related to alertness, attention, vigilance, and problem-solving activities. Theta activity is involved in a decrease in alertness and a lesser processing of information; there are two types of theta activity: (i) low alertness (which leads to sleep) and (ii) selective cognitive inactivity due to automated processing. Delta signal is more prominent during sleep, relaxation or fatigue, and it diminishes with increasing age [101].

The artifacts product of the movement become problematic in all of the physiological measures. However, these are especially problematic with EEG since electromagnetic activity has a very low level of energy. One last major problem with EEG as a measurement signal is the difficult interpretation of the obtained data [88].

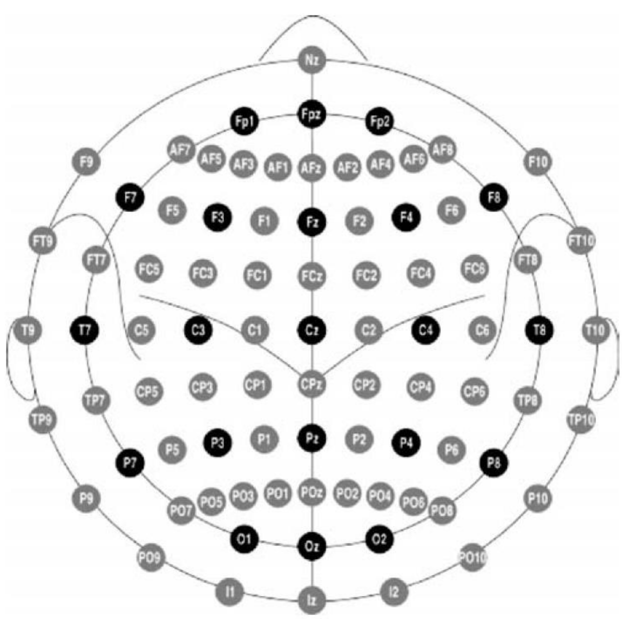

Figure 5. Position of electrodes and labels in the international 10-20 system [88]

EEG measures the oscillations of inhibitory and exciting brain activity, that is, the EEG allows us to record electrical activity in the scalp that is related to brain activity. Generally, the use of the amplitude and frequency of the signal distinguishes the brain activity. There are a variety of different measurement devices available for this type of physiological signal, ranging from more sophisticated medical grade head configurations with high-density electrode arrays (from 32 to 256 electrodes) to simpler devices that have fewer electrodes and less spatial precision (Figure 5) [100]. Table 2 shows the advantages and disadvantages of measuring EEG.

Table 2. EEG's advantages and disadvantages [89]

Advantages	Disadvantages
Major time resolution	Low space resolution
Presents quantitative data	Gel-based caps and conductivity
Different analyses with the same data	Several movement artifacts
Deep cognitive insights	Difficult to interpret

- **<u>Peripheral Nervous System</u>**

The PNS includes all the nerve cells outside the CNS and connects the CNS with the rest of the body. The PNS transmits most of the physical sensations. Consequently, the surface of the skin is the location from which physiological sensors commonly monitor their signal. The PNS is further divided into the somatic and autonomic nervous system. The Somatic Nervous System (SNS) regulates the body activity that we have under conscious control, such as deliberate muscular activity.

The Autonomic Nervous System (ANS) controls the unconscious and visceral responses. Within the ANS, we have the sympathetic nervous system, which triggers fight or flight reactions in emergency situations, and the parasympathetic nervous system, which controls relaxation, rest and digestion. It is important to keep those two systems (SNS and ANS) in mind when measuring psychological events such as cognition with physiological sensors [101]. The most common techniques to monitor the PNS activity are:

a. Electrodermal System

Electrodermal Activity (EDA) has been one of the most used response systems in the history of psychophysiology. The application of EDA measurements to a wide variety of situations is because of the relative ease of measurement and quantification, combined with their sensitivity to states and psychological processes [103].

When measuring the conductance level of the skin over time, we refer to this as EDA measure. When measuring a direct response to a stimulus, we use the term galvanic skin response. In any case, EDA measures the changes in the passive electrical conductivity of the skin, in relation to the increases or decreases in the activity of the sweat glands [104].

Since we measure the differences in conductivity, we only need two electrodes for the measurement, which makes EDA a very easy physiological measure to prepare and apply. EDA is also very easy to interpret since it almost has a one-to-one relationship with physical arousal.

One benefit from EDA measure is making the analysis at a macro level (in larger pieces of game time) or at a micro level (related to events). When analyzing the response to a direct event, EDA is a relatively noisy signal and has some latency in response to a stimulus [101].

EDA is a sensitive peripheral index of SNS activity and a useful psychophysiological tool with wide applicability. Social and behavioral scientists have discovered that tonic EDA is useful for investigating general states of alertness while phasic EDA is useful for studying multifaceted attentional processes as well as individual differences [103]. Table 3 shows the advantages and disadvantages of the EDA measurement.

Table 3. EDA's advantages and disadvantages [89]

Advantages	Disadvantages
Low-cost hardware	Noisy signal.
Ease of measurement.	Great individual variation.
Ease of analysis.	Fluctuations in the baseline and the response.
Less intrusive than other biosensors.	Hardware deterioration over time.

b. Cardiovascular Physiology

The cardiovascular system is essential for life and has been a central focus of psychophysiological research for several reasons. First, at least some of its parameters, such as heart rate and blood pressure, are easily observed and quantified.

Second, the cardiovascular system is a rich and intricate physiological system with multiple regulatory subsystems subject to central and peripheral autonomic controls. Consequently, it is highly sensitive to neurobehavioral processes. Finally, the complexity of the cardiovascular system makes it susceptible to a variety of disorders affected by psychological factors such as stress [105].

There are many cardiovascular measures available for physiological evaluation, and all of them are related to the heart rate, its changes and how this influences the physiological state of the person. The most common measures are Electrocardiography (ECG), Heart Rate (HR), Pulse Interval (PI), Heart Rate Variability (HRV), Blood Volume Pulse (BVP) and Blood Pressure (BP). However, blood pressure is not a measure in real time; for all other relevant measures, we require physiological electrodes to collect the necessary data.

The ECG measures the electrical activity caused by the heart that pumps blood and is usually measured with three electrodes or wires (positive, negative and neutral), and are usually applied to the upper part of the body.

Heart rate is the number of heartbeats per unit of time (usually measured in beats per minute). The PI, or the time between the heartbeats, is also of interest. If the PI decreases, the HR obviously increases, and it is related to a greater processing of information and emotional activation. Then, the IP and the HR are two related measures. However, the variability of the heart rate is a more complicated measure (with a complex analysis procedure) since it depends on the spectral analysis of the synodal arrhythmia. In simple language, we are analyzing changes in frequency and period in PI over time.

In general, we must bear in mind that most cardiovascular measures are intrusive and not accurate. They are affected by many variables, such as physical activity [101]. Table 4 shows some advantages and disadvantages of cardiovascular measures.

Table 4. Advantages and disadvantages of cardiovascular measures [89]

Advantages	Disadvantages
The heart rate is easy to measure.	Intrusive sensors.
The hardware for measuring heart rate is low-cost.	Affected by many non-controllable variables.

The cardiac measures are established and prominent.	The variability of the heart rate has a complex procedure of analysis.

c. Respiratory System

It is often overlooked in psychophysiology since the respiratory system is remarkably complicated and sensitive to a variety of psychological variables. Due to the importance of respiratory function for well-being and other physiological systems, it is widely measured in medical settings. These measures are standardized and can have a variety of uses for psychophysiologists interested in this system. Some of the most commonly used measures are presented below.

The Respiratory Rate (RR) is the number of respiratory cycles that occur in one minute. This value ranges from approximately 12 to 15 breaths per minute under resting conditions. The Tidal Volume (TV) is the normal volume of air inhaled after an exhalation. It is very variable however, it averages around 0.5 liters. The Vital Capacity (VC) is the volume of a complete expiration; this metric depends on the size of the lungs, the elasticity, the integrity of the respiratory tract and other parameters, and so, it is variable among subjects. The values can vary from two to seven liters for adults. Residual Volume (RV) is the volume that remains in the lungs after maximum exhalation.

Like other measures depending on the size of the chest and lungs, this one tends to be highly variable, from 1.4 to 1.9 liters in larger subjects. Compliance (C) is a measure of the elasticity of the thoracic area and it is applied to the Lungs (LC), the Chest Wall (CWC), or the Thorax (TC). Conceptually it is the resistance to change in response to pressure.

Because adequate ventilation is necessary for all physiological systems to work, it is not surprising that the alteration of breathing produces a cascade of changes throughout the body. These changes affect the complete functioning of the organism, including physiological and psychological effects.

The conductance of the skin and other forms of EDA are dramatically influenced by respiration. Asking subjects to take a deep breath is often used to validate the responsiveness and integrity of the recording system. Despite the sensitivity of EDA to respiratory changes, the removal of artifacts due to breathing is difficult.

The effects of the respiratory cycle on the heart are well known to psychophysiologists. The period of the heart, the time between successive beats, lengthens during the exhalations, which produces fewer beats per minute. During inhalation, the heart period is shorter and, consequently, the heart rate seems to increase during this phase of breathing. This oscillatory interaction between the heart and the respiratory system is known as Respiratory Sinus Arrhythmia (RSA) and results from the influence of a variety of physiological systems.

Like the actions of the heart, breathing is clearly a critical process that sustains life. As might be expected from any process so important for survival, it is controlled in multiple ways and affects many body systems. Likewise, it receives a lot of information for its regulation. Traditionally, psychophysiology has concentrated on other dependent variables. When breathing is measured, it is often used as a covariate or a control variable for the variables of primary interest [106]. We present the advantages and disadvantages Table 5

Table 5. Advantages and disadvantages of measuring the respiratory system [76]

Advantages	Disadvantages
Ease to interpret	Affected by physical activity
CWC and TC are ease to measure	Inhibits movement
Low-cost sensors for CWC/TC	The most intrusive sensors

d. Electromyography - EMG

Electromyography (EMG) measures whether the muscles are active or not. For this reason, the measure of the muscle only needs one EMG electrode [107]. When a muscle is flexed in the body, it produces a difference in electrical activity or isometric tension. The EMG tries to measure the activation of PNS, in contrast to other techniques focused on the activation of the CNS.

EMG is a measure of great interest to interact with computers in a more natural way since most of the muscles control is direct. The most frequent use of this technique in measuring the activation of specific facial muscles responsible for showing our positive or negative reactions to a specific stimulus [27]. In particular, research on play physiology has focused on using forehead muscles to indicate negative emotions and cheek muscles to indicate positive emotions or even fun and fluency in a game [108].

The risk of facial EMG is the presence of considerable unnecessary muscle activity. For example, experimental participants cannot chew gum, laugh, or talk during facial EMG, because this introduces large artifacts into the data. Careful processing of the signal in the EMG data must be performed before interpreting them.

In most cases, when electrodes are on the face, people may feel the need to cause more pronounced muscle movements and this could lead to unnatural signals, which would hinder the interpretation of the data [101]. EMG is a non-invasive and accurate way to measure muscle contraction continuously in situations where observation is too imprecise, uncomfortable or costly [109]. We summarize the advantages and disadvantages of this signal in Table 6.

Table 6. EMG's advantages and disadvantages [89]

Advantages	Disadvantages
Great time resolution	Muscle and movement interference
Easy signal analysis	Electrode placement in the face
More precision than face cameras	Difficult to get a natural measurement
Best way to measure emotion	Expensive

4.2.2. Selection of the Physiological Signal

After defining the different kinds of signals, the selection process of the physiological signal that we use in the evaluation of the UX has two components. First, we have the definition of the analysis criteria according to the characteristics of the signals described in the previous subsection. Second, we need some requirements to meet for performing the objective measurement of the physiological signals in the most natural way possible, so that there will be no risk of bias when obtaining the necessary data for the processing of the information.

- **Analysis Criteria**

a. Signal and/or Output

When describing the procedures used to analyze psychophysiological data, it is useful to distinguish between the different stages of the analysis. The first stage is the improvement of the signal and the elimination of observations that are artifacts or can be considered outliers. The second stage involves data reduction: most psychophysiological experiments include a large number of observations per subject and, while this provides datasets that are rich in information content, it also increases the likelihood of spurious or noisy observations. The third stage of data processing is statistical analysis, which may include hypothesis testing, model fitting, parameter estimation, etc. [110].

This analysis criterion is greatly important since the processing of the signal does not constitute a contribution to the research work we carried out. Thus, the simpler this process is, the physiological signal is better punctuated for the instantiation of the evaluation method.

b. Associated Psychological Events

Psychological responses or events are the phenomena related to experience, cognition, emotion, and the behavior of organisms in the physical and social environment [88]. Psychophysiology emphasizes its investigation in the map of the relationships between physiological responses and psychological events (Section 3.3.2. Conceptual Framework). For the criterion of analysis, the ideal relationships between physiological signals and their psychological counterpart are the one-to-one relationship and the many-to-one relationship.

c. Sensor Intrusiveness

This criterion of analysis is very important due to the type of target population of this research: children with SLD. Ideally, the sensors capable of measuring the physiological signals considered cannot feel like a "burden" for the child, on the contrary, the fact of having a device like these should not generate discomfort, tension or any another artifact that acts as a noise when obtaining the signals.

We expect from the sensor to be as unobtrusive as possible so that it does not affect the child's behavior and natural reactions when interacting with the system.

d. Literature Information

The information found in the specialized literature is of vital importance because there are discoveries that others have already covered and have made known to everyone. It is important to be clear that the more information in the literature is about the implementation of physiological signals in the evaluation of UX, the better we score this selection criterion.

- **Analysis of Physiological Signals**

Table 7 shows the scores from one to ten, being one as far away from the ideal established and ten as close or ideal. We present the total score average according to the established criteria of analysis for each of the physiological signals.

Table 7. Scores of physiological signals.

Physiological signal	Criterion A	Criterion B	Criterion C	Criterion D	Total
Electroencephalography	5	4	3	9	5.25
Electrodermal activity	8	7	8	9	8
Cardiovascular measures	7	5	5	8	6.25
Respiratory system	8	3	6	7	6
Electromyography	8	9	4	8	7.25

According to the results of the previous table, the physiological signal we selected for the implementation of our procedure is the EDA. The selection of this signal involves the measurement of psychological events through the interpretation of the stimulation of cognitive performance. Moreover, several studies consider the EDA as the simplest of the signals and yet providing multiple resources to map the psychological events such as emotions or the cognition [54], [60], [111]–[113].

4.2.3. The Procedure

By following the same illustrative structure of the method (Figure 4) and keeping in mind the target users of the HapHop-Physio application, we consider our procedure approach as simplistic: the EDA becomes the convergence point between the player interaction experience of each child with SLD and the measurement of the cognitive skills such as memory or attention. The proposed procedure keeps the same section division and element division from the method (Figure 6).

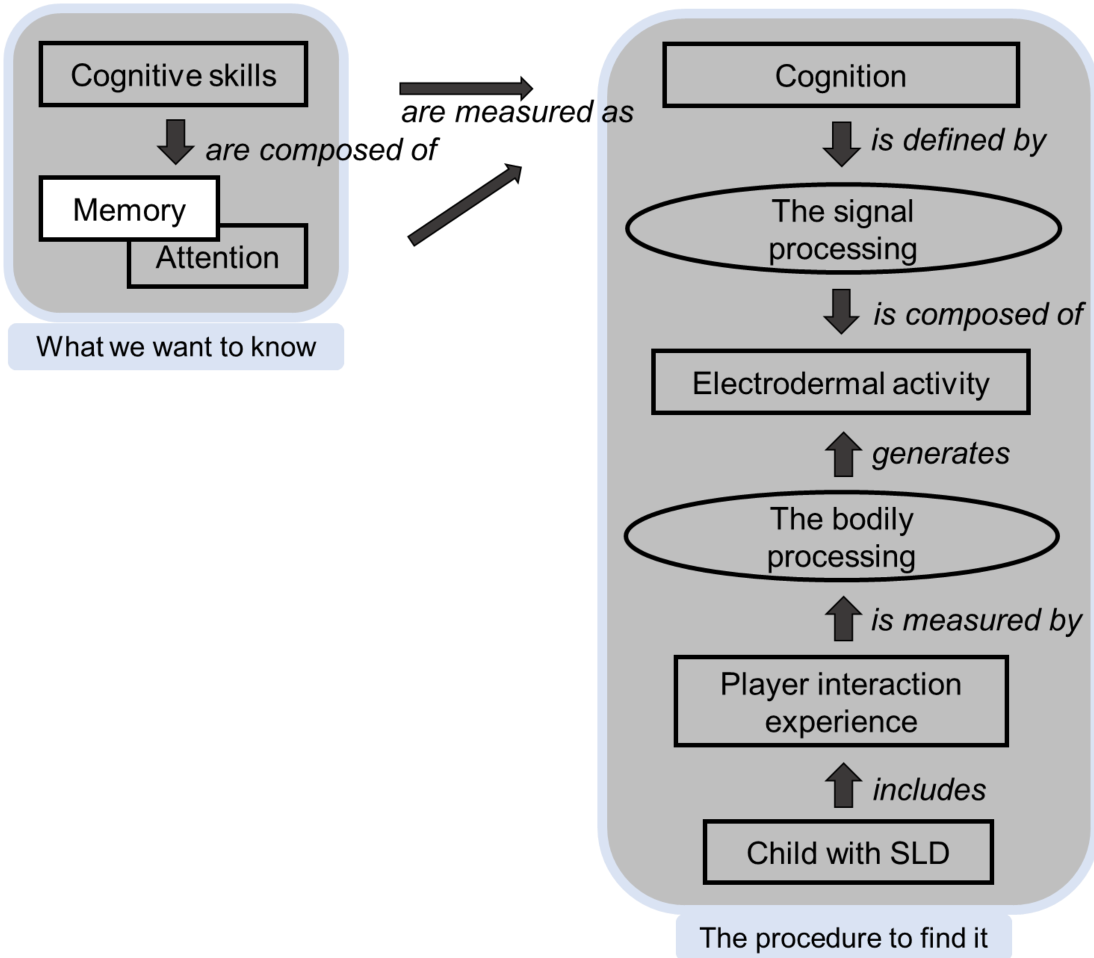

Figure 6. Objective procedure for the UX evaluation, instantiated from the objective method.

In this objective procedure, the player interaction experience of the child with SLD includes the measurable properties associated with the cognition. The bodily processing measures such properties. This is the first of the processes we set up within the method. In the next section, we detail some of the processes occurring during the data collection of EDA. The EDA measures the results of the bodily processing.
The signal processing maps the relationships between EDA signals and the cognitive skills. A convenient way to understand the potential of these relations is to consider a many-to-one relationship, where one body signal is associated with the cognitive activities performed by the user. The result of the signal processing is obtaining the cognition. We describe this second process from the method in the next section. In this way, the cognitive skills become the quantification of the cognition.

By obtaining the final measurement procedure, we conclude the research and development of the conceptual basis of this master's thesis. In the following section, and for the rest of the document, we frame (as a complementary piece for this work) the research and development exercise applied within the data collection to obtain the validation of what we have proposed for the contribution in the telematics engineering area.

4.3. Procedure Application

For the application of the measurement procedure instantiated from the objective method, we define three important approaches to explain them in detail. The first one is to find the physiological signal (point of convergence in the method). To achieve this, we detail the selection of the physiological sensor to measure the signal. The second approach is about the understanding of the bodily processing (the first process within the method) with the aim of having a better insight into the course of action to obtain the signal. Finally, the third approach deals with the signal processing (the second process within the method) to finally obtain the information about the UX that we are looking for in this research work.

4.3.1. Tool Selection (Technology)

To measure physiological signals, we must draw on to the use of physiological sensors. To select the sensor to use, we must have two considerations: (i) what we want to know about the UX, and (ii) how the player interaction experience is affected by the sensor utilization [76]. We already resolved the first consideration through the conceptual framework, the method, and the procedure. Regarding the second consideration, during the search of the devices we took into account, in addition to the aspect of intrusiveness, the following aspects:

- The use of sensors should not be invasive/intrusive in the user's natural behavior.
- Using a sensor should not be expensive to allow its acquisition and use in the field of academia.

- Have the ease of use and interpretation, and that this is allowed by the company that manufactured the device.
- Sensors especially design for children.

Based on these considerations, we mapped four possible options of sensors, devices, and platforms that were built and developed to obtain the EDA.

- **Biosignalsplux's Electrodermal Activity Professional Sensor**

This EDA sensor is capable of accurately measuring the activity of the skin with high sensitivity in a miniaturized form factor. The design of the low noise signal conditioning and amplification circuit provides optimal performance in detecting even the weakest electrodermal skin response events[3].

Features:
Range: 0-13 μS
Bandwidth: 0-3 Hz
Input impedance:> 1 GOhm
Common Mode Rejection Ratio (CMRR): 100dB

- **Movisens's EDA and Activity Sensor**

The 'edaMove' is a psychophysiological ambulatory measurement system that can detect and measure EDA. The sensor acquires the raw data of the EDA and the 3D acceleration of a subject for up to 2 weeks. From these data, secondary parameters such as skin conductance level (SCL), skin conductance responses (SCR) and intensity of activity are calculated with the software 'Movisens DataAnalyzer'.

With the precise technology of 3D acceleration capture coming in all our sensors, 'edaMove' is optimized for research applications. The ability to capture movement, barometric pressure, and temperature allows a more accurate analysis of the data. These parameters allow the clear identification of measurement devices that normally

[3] https://store.plux.info/professional-sensors/280-electrodermal-activity-eda-820201202.html

make it difficult to evaluate EDA data in an outpatient setting. The sensor is used with a wristband and comes with multiple-use sintered electrodes[4].

- **Empatica's E4 wristband**

The E4 wristband is a portable research device that offers real-time physiological data acquisition and software for in-depth analysis and visualization. Some of its characteristics are: (i) discrete monitoring: recordings in the laboratory or at home without complications, (ii) observation of clinical quality: obtaining accurate and accurate physiological data, (iii) data anywhere: easy access and analysis of the raw data with the platform in the secure cloud, and (iv) development of own applications: using the mobile API and Android SDK to build an application and access E4 data in real time.

The E4 wristband is equipped with sensors designed to collect high-quality data:

- Photoplethysmography sensor: measures the pulse of blood volume, from which the variability of the heart rate is derived.
- 3-axis accelerometer: captures activity based on movement.
- Event mark button: marks events and links them to physiological signals.
- EDA sensor: measures constantly fluctuating changes in certain electrical properties of the skin.
- Infrared thermopile: reads the peripheral temperature of the skin.
- The internal clock in real time: temporary resolution of 0.2 seconds in transmission mode.

The E4 wristband has an internal memory that allows to record up to 36 hours with a synchronization resolution of five seconds. An ideal solution for longitudinal studies[5].

- **MySignal's GSR Galvanic Skin Response Sensor PRO**

This sensor measures the electrical conductance of the skin, which varies with its moisture level. This is of interest because the sweat glands are controlled by the sympathetic nervous system, so the moments of great emotion change the electrical

[4] https://www.movisens.com/en/products/eda-sensor/
[5] https://www.empatica.com/research/e4/

resistance of the skin. 'MySignals' is a development platform for medical devices and eHealth applications. MySignals is used to develop any eHealth website or even add own sensors to build new medical devices.

The complete 'MySignals HW' kit includes 17 sensors that can be used to monitor more than 20 biometric parameters. All data collected by 'MySignals' is encrypted and sent to the user's private account in 'Libelium' Cloud via WiFi or Bluetooth. The data is displayed on a tablet or smartphone with Android or iPhone applications. This kit needs Arduino Uno[6].

- **Selection of the sensor**

According to the services and sensors offered, the aspect of intrusiveness, and other aspects considered in the description of each of the devices that we found, we decided to select the E4 wristband from Empatica Inc., a spin-off from the Massachusetts Institute of Technology - MIT. We took this decision because of the advantages presented in this device as the access to raw data, the ease of use for the user due to the wristband shape, the cost, and the fact that its design is usable for children, in regard the other options. Besides this sensor is classified as a medical device for the European Union under the CE Certification N° 1876.

4.3.2. The Bodily Processing

EDA is the general term to define the autonomic changes in the electrical properties of the skin. The most widely studied property is the conductance of the skin, which is quantified by applying an electrical potential between two contact points with the skin and measuring the resulting current flow between them [114]. We measure EDA at the surface of the skin; it arises when the skin receives innervating signals from the brain. The brain sends signals to the skin to increase the sweating level when experiencing emotional activation, cognitive workload, or physical exertion. There might not be any sweat on the surface of the skin, however, the pores begin to fill below the surface in a measurably significant way [115].

[6] https://www.cooking-hacks.com/mysignals-hw-ehealth-medical-biometric-arduino-complete-kit/

There are two main components when performing an EDA measurement. The first component is the tonic level, related to the components of slower action and the background characteristics of the signal (general level, slow climbing, and slow declinations over time). The most common measure of this component is the Skin Conductance Level (SCL); the changes in the SCL reflect the general changes in autonomous activation. The second component is the phasic response, referred to the elements of the signal changing more rapidly such as the Skin Conductance Response (SCR) [114]. In Figure 7 and in Figure 8, we present the division of the EDA components and the differences in the signal between the SCL and the SCR.

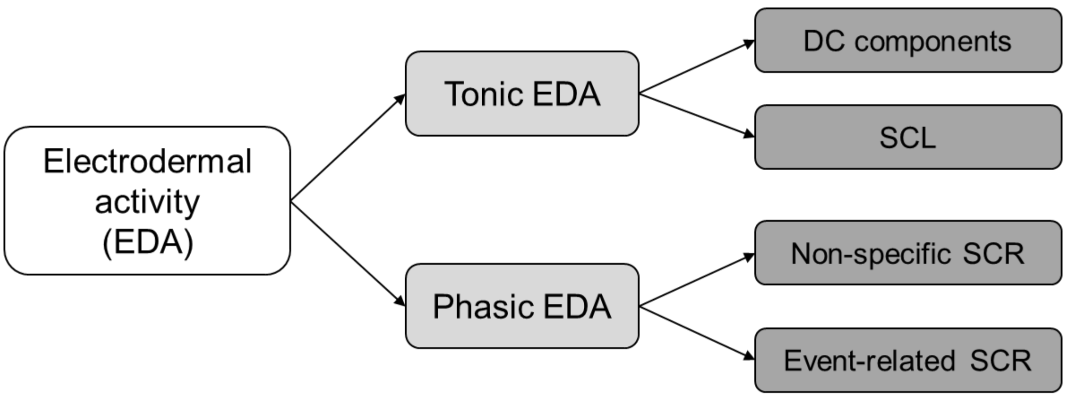

Figure 7. EDA components [114]

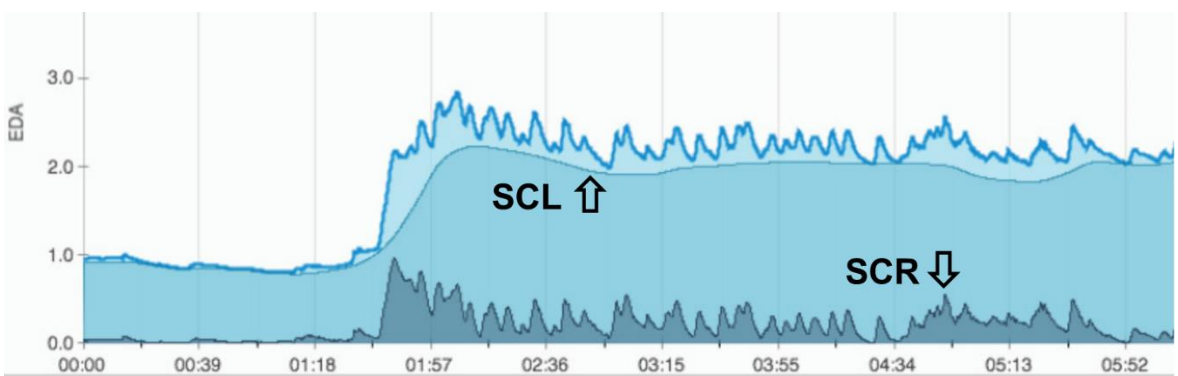

Figure 8. EDA signal, SCL signal, and SCR signal [116]

Phasic EDA measurements associate short-term events elicited one to five seconds after in the presence of discrete environmental stimuli. Phasic changes usually show up as abrupt increases or peaks in the skin conductance. Tonic EDA is generally considered as the level of skin conductance, relatively stable, and associated with gradual changes. The tonic level slowly varies over time in an individual depending upon his/her psychological state and associated with both cognitive and emotional arousal [115], [116]. Researchers have found that tonic EDA is useful for investigating general states of arousal and alertness, whereas phasic EDA is useful for studying multifaceted attentional processes [103].

Although EDA varies widely among subjects, it is relatively stable within each subject. Different psychological states associate the changes in EDA within the subject. Consequently, the analysis of EDA helps in understanding the response of children, particularly the ones affected by SLD, to various environmental conditions [116].

The general recommendation to measure the EDA signal dictates to record it in the palmar surface of the non-dominant hand. This recommendation is for optimal recording conditions based on a maximum signal-to-noise ratio under laboratory conditions. This is not practical for studies in real environments with long recording periods. For ambulatory use in the real world, the Empatica E4 device optimizes the recording activity of the signal. Obtaining the EDA from the wrist is much more practical, less intrusive, and more natural in the common behavior of the user. In practice, the EDA recording of the wrist has a favorable signal-to-noise ratio in real-world conditions when people must use their hands [115].

- **Collection of the EDA Signals**

We use the E4 device to collect the EDA signals from children with SLD. The device is as easy as wearing a watch (Figure 9). There are some recommendations to follow to control the quality of the EDA signal:

- Use the E4 with the case on the upper part of the wrist; use it comfortably; however, tight enough for preventing it from moving around.
- Register the EDA on the non-dominant side of the user to minimize motion artifacts. Nevertheless, this depends on the final purposes of the measurement.
- Place the E4 button on the same side as the thumb or on the other side, any orientation works well.
- Put the EDA electrodes inside the wrist. Optionally, we can align them with the ring finger, however, this is not mandatory.

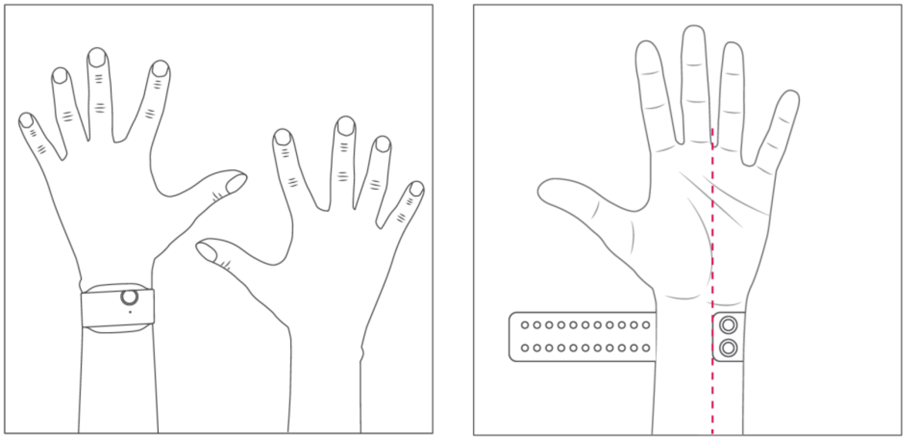

Figure 9. Using the E4 device as wearable [117]

The Empatica E4 wristband is a portable wireless device designed for comfortable, continuous and in real-time data acquisition, which makes it the ideal device according to our requirements. E4 uses the exosomatic methodology (the use of external current) to acquire the EDA signals [118]. We use the E4 in the recording mode within its internal memory. The goal of the E4 is to wear it on the wrist for the continuous collection of physiological data in clinical settings used for health professionals. We present the technical specifications of E4 device for the collection of EDA signals in the Table 8.

Table 8. E4 technical specifications [117]

Characteristic	Description
Sampling frequency	4 Hz
Resolution	1 digit - 900 pSiemens
Range	0.01 µSiemens – 100 µSiemens
Alternating current with a max peak to peak of 100 µAmps.	8Hz frequency at 100 µSiemens
Electrodes	Snap-on silver (Ag)

The recording mode of the E4 allows us to record the data and analyze it later. We use the E4 manager to synchronize the session data of each child to the web dashboard, configure the E4 clock and manage the firmware of the device. In the E4 web dashboard, we:

- View, manage and download all sessions recorded by E4 (Figure 10).

- Observe the details of the session by time series for each type of signal (EDA, BVP, HR, TEMP, and ACC) with overlapping event mark labels.
- Download the raw data in CSV format to facilitate processing and analysis in other applications.

Having the previous description of the approach to the bodily processing for obtaining the raw EDA signal, and the inclusion of practical knowledge for acquiring the signal with the E4 device, we conclude the explanation of the first process presented in the measurement procedure. It is valuable to highlight that the information flow keeps its track on the implemented procedure and the objective method.

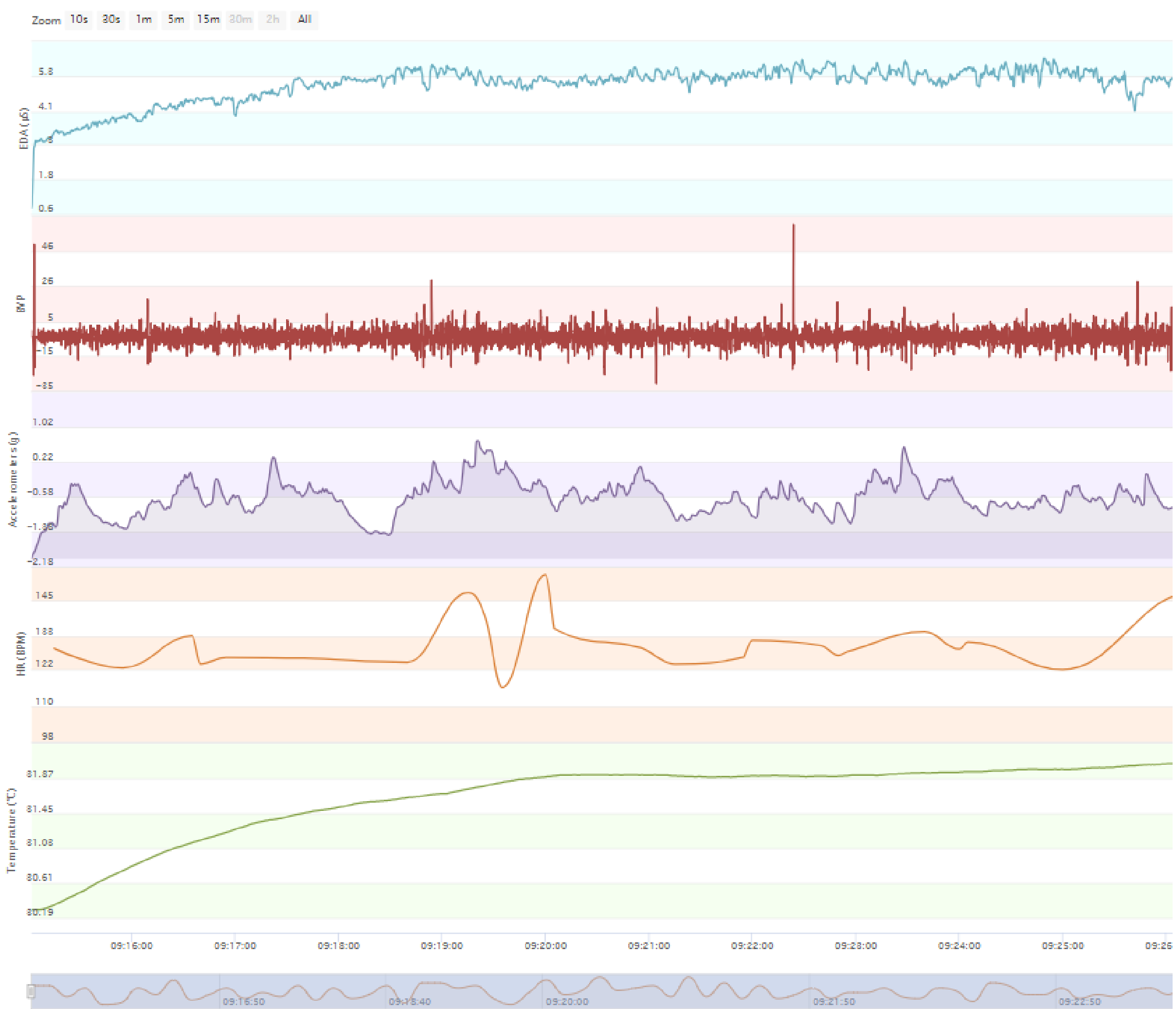

Figure 10. Dashboard for managing the E4 sessions.

4.3.3. The Signal Processing

The second process of the method (psychophysiology) instantiated in the procedure as the signal processing links the description about the relationships that psychophysiology handles to map from the physiological signals to the psychological events (3.3.2. Conceptual Framework). Explicitly, we know EDA is a signal with a many-to-one relation. As we have stated before, researchers relate this signal to the emotion recognition [118]–[121], for our particular case, the recognition of cognitive abilities.

The final aim for processing the physiological signal (any signal) lies in obtaining new knowledge from the analysis of the data found, measured in the user during a given interaction. As a common practice, researchers adopt four basic steps for the signal processing [122]:

1. Raw signals are pre-processed to eliminate contamination (noise, external interference, and artifacts).
2. The features of the signal are extracted, which can be statistical, temporary, frequency-related and temporal-frequent. The features number vary between a dozen and hundreds depending on the number and type of signals analyzed.
3. Those features that might not have a significant correlation with the psychological event under evaluation are eliminated. This increases the performance of the classifiers by reducing noise, which allows a better separation of space and improves time and memory efficiency.
4. A classifier is trained to obtain the classification model using the previously selected features [123].

We report more information about the methods used for each one of the steps for signal processing in Table 9, including some references to related works.

We carried out the signal processing set up in the measurement procedure after collecting the EDA signals. We performed the analysis of data in the environment defined for this work: the interaction in a real environment between children with SLD (in a number of rehabilitation sessions) and the HapHop-Physio application. We explain the selected methods for compliance with each step of the signal processing in the next chapter.

Table 9. Methods for signal processing [123].

Step	Methods	Related works
Filter the signal	• Segmentation • Discard of initial and end signal bands • Smoothing filters • Low-pass filters • Baseline subtraction • Normalization • Discretization	[112], [118], [124]
Feature extraction	• Rectangular tonic-phasic windows • Moving and sliding features • Transformations • Principal, independent and linear component analysis • Projection pursuit • Nonlinear auto-associative networks • Multidimensional scaling • Self-organizing maps	[112], [125], [126]
Feature selection	• Sequential forward/backward selection • Sequential floating search • Branch-and-bound search • Best individual features • Principal component analysis • Classifiers • Analysis of variance methods	[118]–[120]
Classification	Deterministic and probabilistic classifiers: • Ensemble methods • Decision trees	[118], [126], [127]

	• Support Vector Machine • Nearest-neighbor • Neuronal networks	

Summary

Building the conceptual framework helped in the construction of an objective method of UX evaluation that takes into account most of the measurement needs of researchers and practitioners. From the proposed framework, we obtained an objective method, which we consider as a first step towards the validation of the framework.

At the same time, we obtained the measurement procedure to be used in our context, defined by the interaction between HapHop-Physio and the target users. This procedure guarantees the next step towards the validation of the proposed conceptual framework.

From the procedure, we specified the technology for the recording of the physiological signal selected (E4 wristband to measure the EDA). EDA is one of the signals with wide applicability in the recognition of different psychological states. Once we defined the signal and the device, it was also possible to defined and comprehend (i) the bodily processing for the recording of the signal, and (ii) the processing of the signal data, which eventually lead us to obtain the information we are looking for about the UX.

In the next chapter, we apply in a practical way those definitions, decisions, and approaches of the measurement procedure that we examined in this chapter, to evaluate the proposed method in the context of the HapHop-Physio application and verify all the contributions, made from the telematics engineering area, during this research work.

Chapter 5

Evaluation

Throughout this chapter, we present the evaluation of our conceptual framework and objective method in two different scenarios. The first one is the peer-review validation to evaluate the conceptual contributions from this thesis. The second one consists on the data collection, analysis, results, and relevant information about the experiment. We performed an experimental design with the objective of detecting whether the physiological signal like the EDA allows us to identify a psychological event such as cognition during the use of the HapHop-Physio game for the cognitive rehabilitation of children.

5.1. Peer-Review Validation

With the first validation strategy within our evaluation, we presented the conceptual framework to different experts, scholars and practitioners in an academic configuration (e.g., a conference) to receive comments and feedback. This task aimed to assess if the proposed framework presents a reasonable theory for different academics working on similar issues [16].

Even though this validation strategy is highly subjective and might draw biased results, it is convenient to estimate the opinion of the academy on the proposed framework [128], [129]. Accordingly, we obtained feedback on the paper describing our proposed framework ([97]) through a peer review from the International Program Boards in the HCI International Conference, and (ii) the presentation of our framework and the preliminary version of our method in the mentioned conference. These academic validation approaches offer quality and reduce the risk of bias in this validation.

5.1.1. Peer Review

The peer review system validates academic work, helps to improve the quality of published research, and increases networking possibilities within research communities. Despite criticisms, peer review is still the only widely accepted method for research validation [130].

The HCI International Conference from 2018 received submissions to be peer-reviewed by at least two independent referees from the International Program Boards[7]. The conference has an h5 index of 20, and Springer published the proceedings in the series Lecture Notes in Computer Science (LNCS), classified as Q2 in SJR and A2 in Publindex. The paper was sent to the "Design, User Experience and Usability" thematic area of the conference.

From the peer-reviewed process, we obtained the following comments:

- C1: "The manuscript needs to be improved, particularly the references to other works".
- C2: "The authors propose a conceptual Framework for measuring UX with objective methods. Although the basic idea of the paper is worthwhile, the approach is imprecise. My criticism concerns the following issues:
    - Although the authors emphasize the importance of objective measurement methods, they do not describe the methods themselves in

[7] http://2018.hci.international/submissions

proper detail and the fail to specify which method should be chosen for which research objective.
  - The framework has not yet been validated: "The validation of the framework should be achieved through peer reviews." Without any validation its usefulness (and usability) remains unclear.
  - The authors evoke the impression that research so far has reduced UX "to a simple aspect of usability". This claim is wrong since the state of the art regards usability e.g., (in terms of pragmatic product qualities) as ONE aspect of UX (and not vice versa).
  - The authors neglect other frameworks and theories on UX, as for example the works of Hassenzahl or Pohlmeyer".
- C3: "The quality of language of the submission is rather poor and requires a thorough revision by a native speaker".

- **Discussion of comments**

For each of the comments and suggestions received, we make some adjustments (Table 10):

Table 10. Adjustments from the peer-reviewed process.

Comment	Adjustments
Comment 1 – C1	We looked for more references in the proceedings of previous versions of the HCI International conference, and in similar conferences such as CHI.
Comment 2 – C2	For each issue, we: • Set up the most used methods for any research objective into categories; however, we clearly stated that our contribution was the definition of a new method. • Explained that the validation of the method depends entirely on the methodology followed to create it. Participating in the conference was the first step towards validating it. An experimental evaluation was declared as future work, which was later performed and reported in the next sections. • Corrected the paragraph about the reduction of UX to a simple aspect of the usability paradigm for a better understanding. However, we did not change the main idea. • Sought, categorized and referenced more conceptual frameworks, especially from the work of Hassenzahl.
Comment 3 – C3	We checked the quality of language through grammatical correction tools.

With the necessary adjustments made, the paper was accepted for publication in the Conference Proceedings and for its presentation. Thus, we continued with the second step in the validation strategy of the conceptual framework: the presentation of the paper.

5.1.2. Paper Presentation

Following the validation guidelines of Jabareen ([16]) for conceptual representations, we orally presented our conceptual framework for UX evaluation in the session *S145: Usability and user experience evaluation methods and techniques*. The session's chair was the Ph.D. in Human-Computer Interaction, Martin Maguire, from the Loughborough University, United Kingdom.

Those attending this session of the conference had the necessary experience in the HCI area to offer us feedback on the work we presented. However, the participation through questions marked the interest of the assistants and the chair towards the application of the conceptual framework and the method in the HapHop-Physio scenario.

We received a couple of observations regarding the presented method:

- Since we focus the construction of the framework and the method for working with children, an attendee suggested me to try working with the "Fun Toolkit" method. However, we already have done it previously. Such work is reported in the paper “HapHop-Physio: A Computer Game to Support Cognitive Therapies in Children” [9]. The results of this work led us to rethink the approaches of evaluation methods for children and, therefore, consider the definition of our own.
- Review the generalization of the method. Within the method, we had an instance. This mistake leads to a poor application of the method by means of a procedure. We later corrected the error found, as it is demonstrated in the previous chapter.

Based on these results, we conclude that the attendees understood the final objective of application of the conceptual framework and the method. However, the validation process was not complete, as we received little feedback.

Finally, regarding the results from the peer-reviewed process and the paper presentation, we have corrected and validated our conceptual framework for the UX evaluation in two different academic configurations, with different scholars and practitioners. Hence, we consider our conceptual framework as “validated by the academy” in a quite reliable qualitative process [129].

5.2. Experimental Validation

To validate our procedure, our method and thus, the conceptual framework, we collected the EDA during the interaction of the children with the active video game for rehabilitation HapHop-Physio. The game is under its validation phase to check its therapeutic effectiveness. We focused part of this validation on determining the UX that children have during the interaction.

Because the UX is a widely disseminated, developed and implemented research area, there are many variables around how we should evaluate and/or measure it. In the case of HapHop-Physio, we used the measurement paradigm of psychophysiology, which relates a physiological response to a psychological event.

In general, the UX approach is based on the recognition of basic emotions based on the circumplex model of affection and on the proven knowledge about the different physiological signals [119], [120], [131]. However, there are other related psychological events such as behavior or cognition.

Based on the literature, EDA could be an indicator of the cognition [116], [132]. In addition, taking into account the rehabilitation goal of HapHop-Physio, we decided to attempt at recognizing cognition from the physiological responses of children during their interaction with the game.

In this section, we describe the real experimentation scenario along with the protocol for collecting data from the EDA. Likewise, we report the process of preparing the data for further analysis, using two approaches: descriptive statistics and machine learning. Finally, we discuss the results obtained through this experiment.

5.2.1. Description of the Experimental Scenario

In the frame of the applied research project called "*HapHop Fisio II - Un Juego Serio Interactivo Basado en Movimiento (Exergame) para el Soporte a Terapias Cognitivas en Niños*" registered under ID 4441 in the Research Information System of the *Universidad del Cauca*, and financed with resources from the *Sistema General de Regalías* of the Colombian government, under the *InnovAcción Cauca* project (ID 3848), we conducted a quasi-experimental pilot study, prospective and longitudinal from April to June, 2018 (8 weeks), in which six children participated. The children selected for this study attend the neurodevelopmental and neurorehabilitation program of the Fisiocenter S.A.S. rehabilitation clinic.

The objective of this study was to determine the changes in the cognitive domains of memory (visual/auditory-verbal) and attention (visual/auditory) after receiving treatment with the HapHop-Physio application. For this study, we considered as inclusion criteria:

- Children aged between 5 and 16 years old.
- Children diagnosed with:
    - Motor development delay, global developmental delay, mild or slight mental retardation, or specific learning disorders.
    - Behavioral problems such as depression, anxiety or emotional factors.
    - Alteration in executive functions such as disorganization and poor planning, amnestic syndrome, and attention deficit hyperactivity disorder.
- Children with an identified deficit of the basic memory and attention processes during the admission evaluation to the neurodevelopmental and neurorehabilitation program.

Together with the specialists of the clinic, we defined some exclusion criteria according to the scope of the therapies carried out with HapHop-Physio. The exclusion criteria were:

- Patients with severe and profound mental retardation diagnose.
- Patients with severe or profound visual or hearing loss.
- Patients who do not have difficulties in memory or attention cognitive skills

- **Study Procedure**

Three phases comprise this study. The first phase is the application of the Children's Neuropsychological Battery (ENI 2) [11]. The main objective of the battery is to determine the presence of cognitive and behavioral changes in individuals who are under suspicion of some type of alteration. This battery measures, in a standardized way, the cognitive profile of children between the ages of 5 to 17 years, by evaluating 12 cognitive domains. However, for this study, only the memory and attention domains were considered with their respective subdomains. From the analysis of the results, the child's profile of execution was obtained to recognize the level of the attention and memory cognitive functions.

The second phase of the study selects the population that, in the application of the battery, presented low average, limit, low or extra low percentile scores in one or more of the cognitive skills evaluated. This population followed a rehabilitation process through the HapHop-Physio therapeutic program.

In the third phase of the study, a revaluation of the cognitive profile of the children was carried out to check the progress and/or decrease of the cognitive difficulties found in the initial evaluation. The fulfillment of the clinical objectives can be considered as achieved when the revaluation of children reaches an average rating, high average or superior since it is an indicator that the patient no longer has difficulties with these skills.

Considering the described study, and to carry out the evaluation of the method proposed in this Master's work, the data of the EDA signal was collected from the six participating children, during a total of 67 recording sessions with the E4 wristband, while the children carried out the therapies under the supervision of the neuro-psycho-pedagogy specialist. Each game session lasted approximately 25 minutes, generating an approximate of 28 hours of EDA signal recordings.

5.2.2. Data Collection Protocol

Both for the pilot study of the project and for the experiment of this research work, we considered the ethical principles established in the Belmont report of 1978, the Helsinki Declaration, and the Resolution No. 8430 of 1993 from the *Ministerio de Salud y Protección Social* in Colombia. We respect the autonomy and integrity of patients who decide to participate voluntarily with prior informed consent in Spanish (Annex B) and we guarantee the confidentiality and privacy of the information collected for the study.

At Fisiocenter S.A.S., we have a separated place completely dedicated to the application of therapies with the HapHop-Physio system. The room has:

- A 32" television to display the game.
- An alphanumeric mat to introduce the answers in the game.
- A computer to run the game's desktop application and to store the data of each therapy session.
- Decoration on the walls according to the graphics style of the game.

Before starting each day of recording, we verify (i) the battery level of the device, and (ii), the synchronization of the E4 timestamp with the computer that stores the data of the game session. For the recording of each session, we followed the same series of steps with each child participating in the study:

1. The child enters the room without his/her guardian and according to the schedule set by the specialist. Most children attended Monday through Friday afternoons, after their school day. On Saturdays, we only had morning sessions.
2. Before starting the therapy sessions, we placed the E4 wristband on the non-dominant side of the child, according to the manufacturer's recommendation. Once we correctly located the E4 on the wrist, we turned on the wristband by pressing its button for two seconds. A green light blinks for approximately 50

seconds, and then it blinks red for a few seconds indicating the start of the signal recording. Finally, the light turns off to save energy.

3. While the EDA recording time begins, the specialist asks the child to remove their shoes and start setting up the game (log in and choose the child's avatar).
4. Once the recording begins, the specialist indicates to the child the first game of the session. The specialist selects the first game in the first session of the child according to the results of the ENI 2 battery. In the following sessions, the game selection depends on the progress of the child. Progress information is on the computer in the room; the specialist searches for it to define the course of the game session.
5. The child performs the therapy for as long as the specialist deems appropriate, according to their professional judgment and the fatigue perceived in the child (approximately between 20 and 30 minutes).
6. Once the specialist indicates the end of the session, we finalized the recording time by pressing the button for two seconds. When this happens, the device saves the session to a file in its internal memory and turns off. As an indicator, the light turns red for one second.

Other considerations regarding the recording of the signals were:

- When changing user, we had to clean all the E4 device parts in contact with the skin. We sanitized by cleaning with cotton moistened in ethyl or isopropyl alcohol.
- The signals of the daily sessions were stored in the internal memory of the device. At the end of the signal capture of the last patient of the day, we used the E4 Manager software to transfer the stored sessions to the web dashboard and thus clean the internal memory of the E4.
- For verifying the information about the readings the sensor made from the children during the interaction time with HapHop-Physio, we recorded each therapy from two different angles: one looking at the child's face from the right or left side and another looking at the screen on which the children watched the game.

In Table 11, there is a summary of the EDA recordings collected by each of the study participants.

Table 11. EDA recording sessions per child

Patient Identifier	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Recovery Week	Total number of sessions
Patient_A	3	0	1	0	3	2	2	1	1	**13**
Patient_B	0	1	2	1	0	1	2	2	2	**11**
Patient_C	2	2	2	1	2	1	2	0	0	**12**
Patient_D	1	1	2	0	2	1	2	1	0	**10**
Patient_E	1	1	2	1	2	1	0	2	0	**10**
Patient_F	1	1	2	1	2	2	1	1	0	**11**
Sessions per week	**8**	**6**	**11**	**4**	**11**	**8**	**9**	**7**	**3**	**67**

Regarding data collection, we had some shortcomings:

- For several reasons, the participants did not regularly meet the appointments for the therapy. Therefore, we had an additional week of recovery. The specialist defined the criterion of addition the number of total sessions of each participant; we required to have at least 50% of the therapies.
- The participants were called to participate two and three times a week, according to the criteria of the specialist. For this reason, we do not have the same number of recordings per child.
- In addition, progress in the game depends on the performance of each child, so not all the children played all the mini-games, nor they reached all the levels.

5.2.3. Data Preparation

The signal of EDA coming from the E4 wristband is a non-stationary signal (i.e., the signal varies over the time) and its amplitude is measured in μSiemens. On average, each recording session lasted about 25 minutes. This duration is estimated on the time assigned to a conventional therapy session for the treatment of children with SLD.

In this way, we proceed to process the EDA signal following the regular procedure found in the literature [123]:

1. Preprocessing of the signal to detect artifacts by movements or poor sensor readings.
2. Extracting features because we are dealing with non-stationary signals; there are different ways of extracting the features.
3. Selecting features to reduce the processing overload on the system.
4. Extraction of the necessary information.

However, the steps mentioned above are still based on the assumption that the signal is labeled, i.e., those stimuli that produced the physiological response are identified, categorized and defined for each of the representations or changes found in the signal. Once we considered cognition as the valuable information that EDA provides for us, it is important to identify the stimuli to which the child was exposed during the treatment session (recorded with the E4 wristband) that may involve some change in the child's physical response.

Bearing this in mind, we needed to segment the EDA signal according to the content of the game addressed during the therapy sessions. The segmentation is not uniform since it depends on the length of time of each game and the levels reached by each child. It is necessary to clarify that besides the EDA signal, we also collected information about the child's performance in each game, in addition to video recordings of the sessions with each child, which is the main base for segmentation and subsequent labeling.

After the signals segmentation, we identify each of the fragments as the stimulus generated from a change in cognitive activity. We process the EDA signal to obtain the features of the signal and subsequently, we label it according to different variables about cognition from the game's database. This kind of information is variable since the difficulty increases with the level, which consumes more time in the stimulus and the child's responses, both physiological and cognitive.

Below, we describe the protocol followed for the segmentation, labeling, and features extraction of the EDA signal for each child.

- **Segmentation**

We proceeded with the segmentation of the EDA signals taking into account the information at hand about the signals and from the videos of the children. We carried out this task as part of the EDA signals pre-processing activity. For purposes of this experiment, we considered the number of participants as enough to generate a "general" model for the examination of the relationship between EDA and cognition, fulfilling the proposal for this master's work. We will record more EDA signals from other participants and use them as the testing set for the generated model in future works.

To segment the signals:

1. We download the CSV file that contains the data of each one of the recorded signals.
2. Based on what we observe in the videos and the information on the HapHop-Physio database, we identify the module, the category, the modality, the level, the score, the duration, the initial time, and the final time of the mini-games played by the children in the treatment sessions.
3. With the time of each mini-game, we segmented the signal from the CSV file, discarding those segments of the signal where the child did not have the kind of cognitive interaction that we are looking for to validate the design of our experiment.

- **Quality of EDA Segments**

To guarantee the quality of the EDA signal data, we based the cleaning process of the signal segments on three rules [133]:

Table 12. Rules for Quality Assessment of EDA data [133]

Rule	Rationale
1. EDA is out of range (not higher than 0.05 μS)	To prevent “floor” artifacts when electrode loses contact with skin, we chose 0.05 μS because it is at an accepted minimum for EDA amplitude
2. EDA changes too quickly (faster than ±10 μS/sec)	To prevent high frequency or “jump” artifacts.
3. EDA data surrounding (within 5 seconds of) invalid portions according to rules 1 and 2.	To take into account for the transition effects that we find close in time to the artifacts according to rules 1 and 2.

We apply the three rules described for each signal segment. Table 13 summarizes the initial number of each signal segment per child, and the final number of signal segments that we use in the experiment. In the end, we obtained a set of 438 EDA signal segments.

Table 13. Number of EDA segments per child

Patient Identifier	Sessions	Segments	Valid segments
Patient_A	13	77	71
Patient_B	11	90	80
Patient_C	12	100	86
Patient_D	10	72	55
Patient_E	10	92	72
Patient_F	11	89	74

- **Labeling**

According to what we have established so far, we are interested in knowing the recognizable aspects of cognition through the EDA. For the task of labeling, we outline five variables related to cognition. We have information about these variables from the stored data in the HapHop-Physio database.

- Session
  A game session is equivalent to a conventional therapy session. Each session has a variable number of mini-games. For our approach, the stimulated domains in the cognitive profile of the child should increase as the sessions of the game increase, according to the specialist in charge of the accompaniment of the sessions. In our case, the addressed domain was the memory, and the auditory-verbal and visual subdomains. This aspects value (session ID) is according to an ascending number of the ordinal category, and is related to the length of the therapy.

- Level
  HapHop-Physio has four levels of depth and five levels of difficulty. For now, we focus on the difficulty levels: the game has five of them (ordinal number one to five). Ideally, this aspect should increase as the number of session increases. We consider the level as an indicator of cognitive workload [132].

- <u>Score</u>

  The score of the mini-games depends on the level, and the neuropsychological evaluation battery (ENI 2 [11]) used in this study defines the scores for each level. For each level to be consistent with the average percentile of cognitive ability, the child should have: (a) three correct answers in the first level, (b) in the second level, five correct answers, (c) the third level has seven correct answers, (d) nine correct answers in the fourth level and, (e) for the fifth level 10 correct answers.

  It is worth noting that the score obtained by the child is the equivalent of the child's memory span. The memory span is the number of items (usually words, letters, or numbers) that a person can retain and perfectly reproduced [134]. That is why we consider it as another aspect of cognition.

- <u>Performance</u>

  We established the performance variable in terms of the score according to the economics factor created for HapHop-Physio. We represent the reward of the game in stars: the player collects one, two or three stars for each completed mini-game. The player never loses stars for responding incorrectly to the challenges of the mini-games and minimally collects one star for each mini-game. Performance maps the scores as shown in Figure 11:

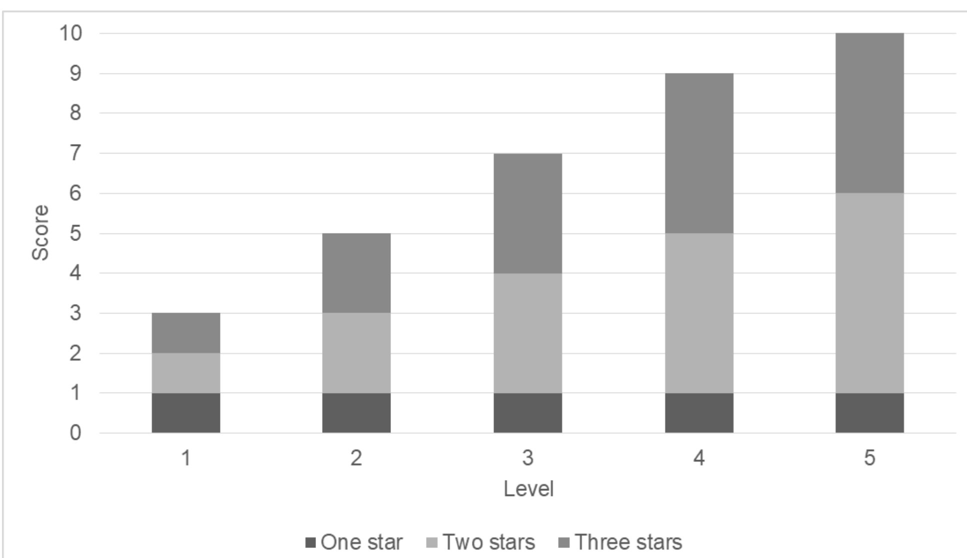

Figure 11. Performance according to Score and Level

- Time

  Time is the variable that measures the duration of each mini-game. For our experiment, this variable is an important aspect of cognition since it gives us information about the speed in the organization and planning of the executive functions of each child [135]. We must realize that low latencies can represent impulsivity in children, however, we cannot say anything about high latencies.

We labeled each signal segment according to the information of these variables in each mini-game. We test the described variables to recognize which of them relates in a suitable way the cognition to the EDA signals.

- **Feature Extraction**

So far, we have approached the first step of the preparation of the signal (segmentation and labeling) to continue with the processing of the signal segments from the steps considered in Table 9 from Section 4.3.3. The first step towards processing is filtering the signal to perform the extraction of features of each segment.

Nevertheless, before going on, we consider relevant to remember that the EDA signal has two components: the phasic component (SCR) and the tonic component (SCL). For the purposes of our experiment, we use the tonic component of the signal. The tonic component represents the change of the signal in time due to the changes generated by the SCR [116]. Given the scope of the information we have at hand (variables), it is not possible to label each of the SCRs produced by the child's interaction with HapHop-Physio. However, we can label the set of changes represented by the SCL.

We filter the EDA segments using a low pass filter specially adapted for the number of samples of each of the segments (e.g., Figure 12). To filter the signal, we employ a convolution process with a sinc (sine cardinal) function in the time domain [136] to obtain the SCL component from the EDA signal (Figure 13).

Since the EDA data are classified as non-stationary signals, multiresolution analysis techniques are appropriate to represent the information contained in this physiological signal. The Wavelet Discrete Transform (WDT) is a widely used technique for multiresolution analysis of EDA data [126]. The WDT consists of comparing the signal of interest with all the possible translations and scales of the same wavelet function, known as the mother Wavelet [137]. Thus, we performed the analysis of our signals (SCL component) to obtain the features of the signal in the wavelet domain. As a result, this domain provides us a more detailed representation of the signal in comparison with the raw time-domain signal.

For the WDT application to each segment, we used the Daubechies wavelet of order 10 (db10) as Wavelet mother [137]. We performed a multiresolution analysis of 10 levels using the Mallat algorithm [138]. As a result, we obtained a total of 11 sets of coefficients (Figure 14), one corresponding to the approximation (scaling coefficients) and 10 corresponding to the details (wavelet coefficients).

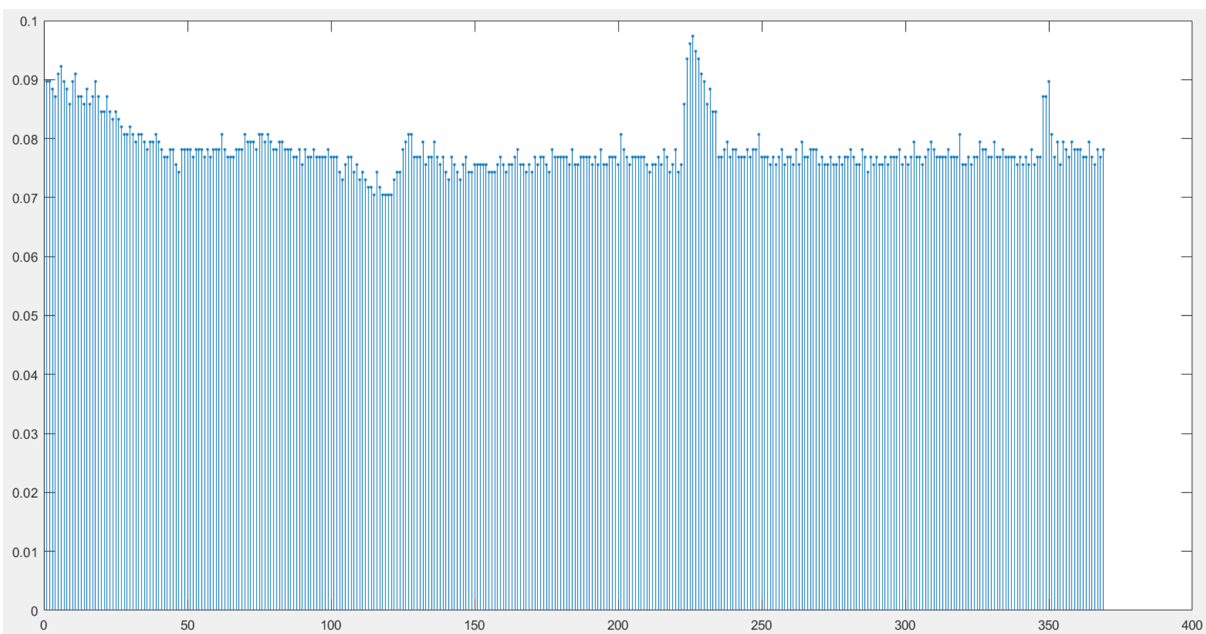

Figure 12. Filtered segment of EDA

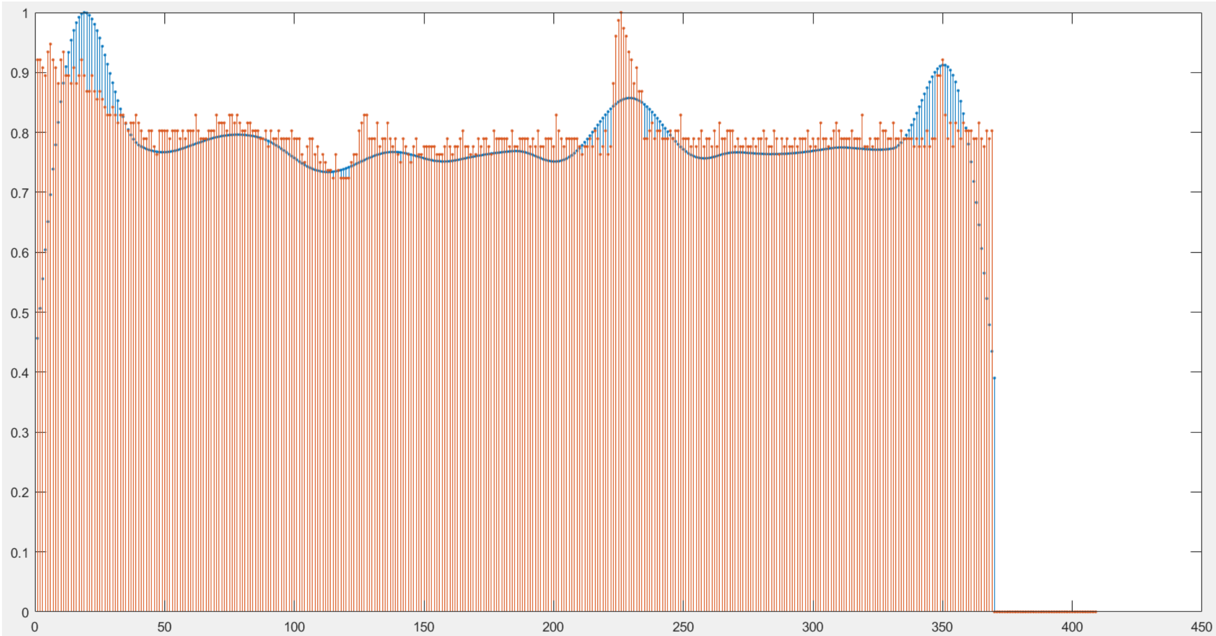

Figure 13. SCL component (in blue) of the segment

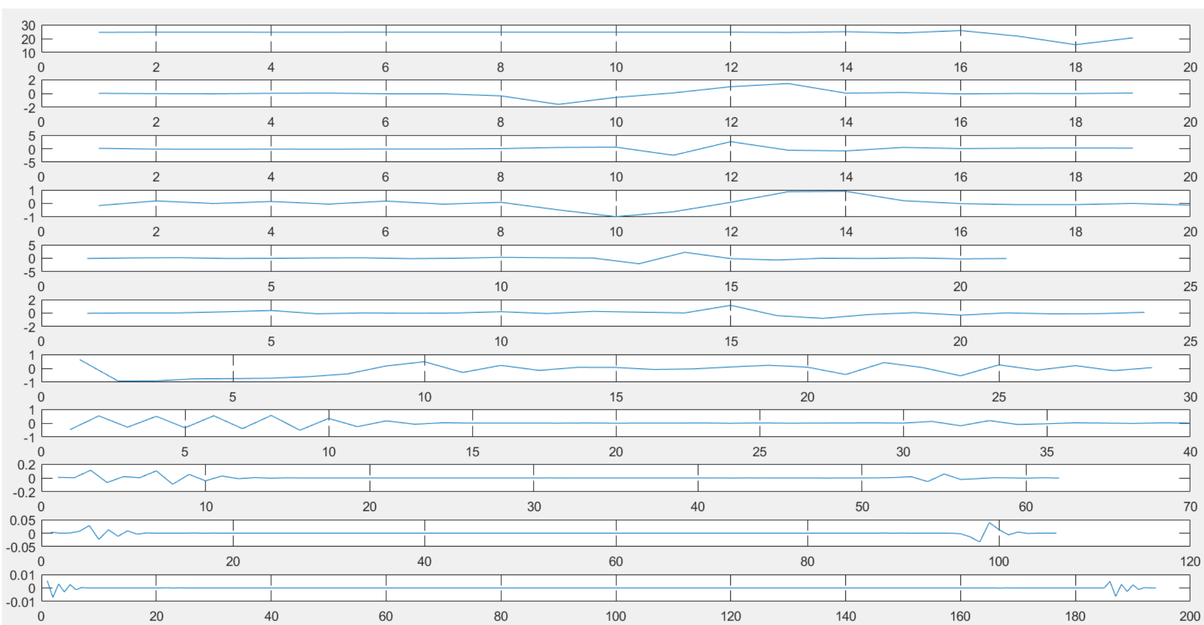

Figure 14. Coefficients in the Wavelet domain.

Proceeding with the feature extraction of the SCL segments (second step in the signal processing) we consider the 10 wavelet coefficients to obtain the equivalent in the wavelet domain of the total amplitude of the signal, the normalized amplitude, and the absolute amplitude from each detail coefficient, the first derivative of the detail coefficient, and the second derivative of the detail coefficient. From each of the previous combinations between the amplitude and the mathematical rate of change of the signals in the wavelet domain, we selected four statistical variables: the mean, the variance, the kurtosis, and the standard deviation. Table 14 shows, in a visual way, the combination of this features.

Table 14. Features extracted in the Wavelet domain

Amplitude variations	Mathematical rates of change	Statistical variables
Total amplitude Normalized amplitude Absolute amplitude	Base signal First derivate Second derivate	Mean Variance Kurtosis Standard deviation

We extracted 360 features from the representation of the SCL segments in the Wavelet domain; 36 features (three variations of amplitude, by three mathematical rates of change, by four statistical variables) for each level of the detail coefficients (10). With the extraction of the features, we have finally generated a data set with 436 samples (instances) and 360 features (attributes).

5.2.4. Statistical Analysis

Summing-up we: (i) segmented the signal and labeled the segments, (ii) filtered the EDA signal of each segment to obtain the SCL, and (iii) analyzed the SCL segments in the Wavelet domain to extract the features and obtain the dataset for subsequent analysis. For selecting the representative features to evaluate each SCL segment, we chose to compare two approaches. The first option is the statistical analysis, which we explain in this section, while as a second option we consider a machine learning approach (Section 5.3.5).

For the statistical analysis and interpretation of the data, the descriptive statistics deals with the graphical presentation and numerical processing of a dataset while the objective of hypothesis testing is to see the possibility to reject a certain null hypothesis based on a sample from some statistical distribution. We used both techniques (descriptive statistics and hypothesis testing) to describe and present aspects of the dataset, such as measures indicating how concentrated or spread out the data is and to better understand the nature of our data [139].

To perform analyses, we defined various independent variables related to one dependent variable: the EDA segments (Figure 15). We divide the independent variables into principal and secondary variables. The main variables are qualitative and are those related to cognition (explained in the Labeling subsection from the Section 5.3.3). The secondary variables are categorical and related to the demographic data of the children participating in the study (Figure 16). The secondary variable called "severity" refers to the cognitive disorder of children. Participants in the study are diagnosed with SLD; however, some participants are also diagnosed with disorders such as mild mental retardation, Down syndrome, and delays in psychomotor development. We classify the severity of the disorder as low, medium and high. For the statistical analysis purposes, we map the severities accordingly to zero, one and two.

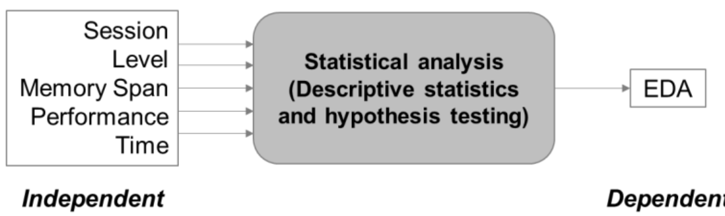

Figure 15. Qualitative independent principal variables

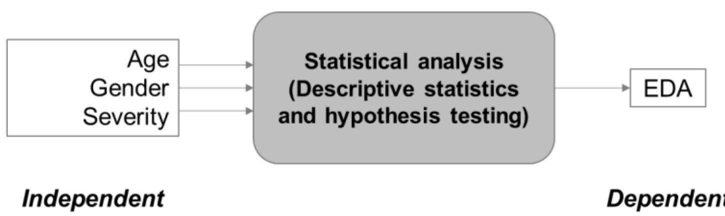

Figure 16. Categorical independent secondary variables

Statistical analyses performed during the verification of this experiment were carried out with a sample of the attributes of the dataset. The selected attributes represent the mean, variance and kurtosis of the equivalent amplitude from the first detail coefficient and its corresponding first derivative.

- **Frequency Descriptive Analysis**

Table 15 and Table 16, and Figure 17 show the results from the frequency descriptive analysis.

Table 15. Descriptive Statistics from Features

		m_d1r	v_d1r	k_d1r	m_d1d1r	v_d1d1r	k_d1d1r
N	Valid	438	438	438	438	438	438
	Missing	0	0	0	0	0	0
Mean		-1.04E-008	1.03E-006	55.802	-2.81E-005	3.50E-006	51.304
Std. Deviation		2.42E-7	7.44E-007	68.344	1.97E-5	2.51E-006	62.562
Variance		.000	.000	4671.004	.000	.000	3914.062
Skewness		3.289	.905	4.000	-.822	.908	3.918
Std. Error Skew.		.117	.117	.117	.117	.117	.117
Kurtosis		24.414	.422	26.255	.283	.432	24.452
Std. Error Kurt.		.233	.233	.233	.233	.233	.233
Minimum		-8.98E-007	1.0E-008	8.573	-9.56E-005	3.5E-008	7.886
Maximum		2.09E-006	3.7E-006	736.47	-3.33E-007	1.2E-005	653.735

Regarding the first attribute of the signal corresponding to the mean of the total amplitude from the first detail coefficient (m_d1r): the mean is -1.04E-008, with a standard deviation of 2.42E-7; it has a zero variance, so we conclude that data are grouped around the mean. It has a positive asymmetry and is leptokurtic. The minimum of the variable is -8.98E-007 and its maximum is 2.09E-006.

The second attribute of the signal corresponds to the variance of the total amplitude from the first detail coefficient (v_d1r): its mean is 1.03E-006, with a standard deviation of 7.44E-007; it has a zero variance, so its data have a homogeneous distribution. It has a positive asymmetry and a low coefficient of kurtosis. The minimum of the variable is -1.0E-008 and its maximum is 3.7E-006. The kurtosis attribute of the signal representation (k_d1r) has a mean of 55.802, with a standard deviation of 68.344. The variance is 4671.004, meaning the set of its data is not homogeneous. It has a positive asymmetry and is leptokurtic. The minimum of the variable is 8.573 and its maximum is 736.47.

The behavior of the next three attributes corresponding to the first derivative of the first detail coefficient (m_d1d1r, v_d1d1r, k_d1d1r) is similar to the previous ones (Figure 17, Figure 18, and Figure 19): the mean of the mean is low and negative, it is positive for the variance and it is high in the kurtosis. Only the mean and the variance have zero variance, while the variance of the kurtosis is 3914.06. Thus we conclude that most of the data are constant since the mean is low as well as the variance; this indicates that there are not many changes in the signal. However, the high values of kurtosis indicate that there are segments of high variability.

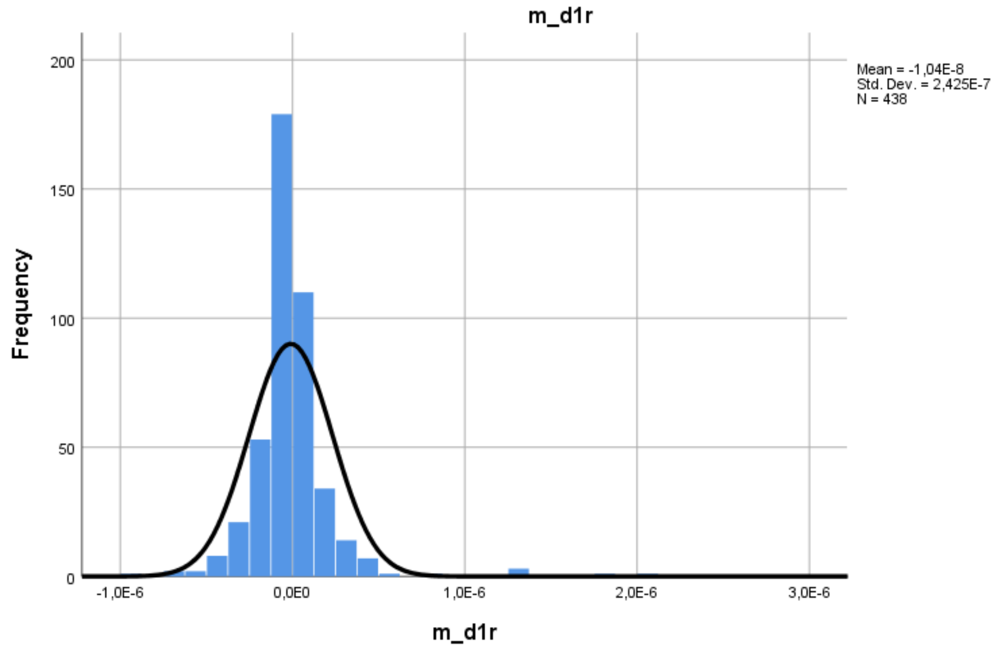

Figure 17. Frequencies of m_d1r

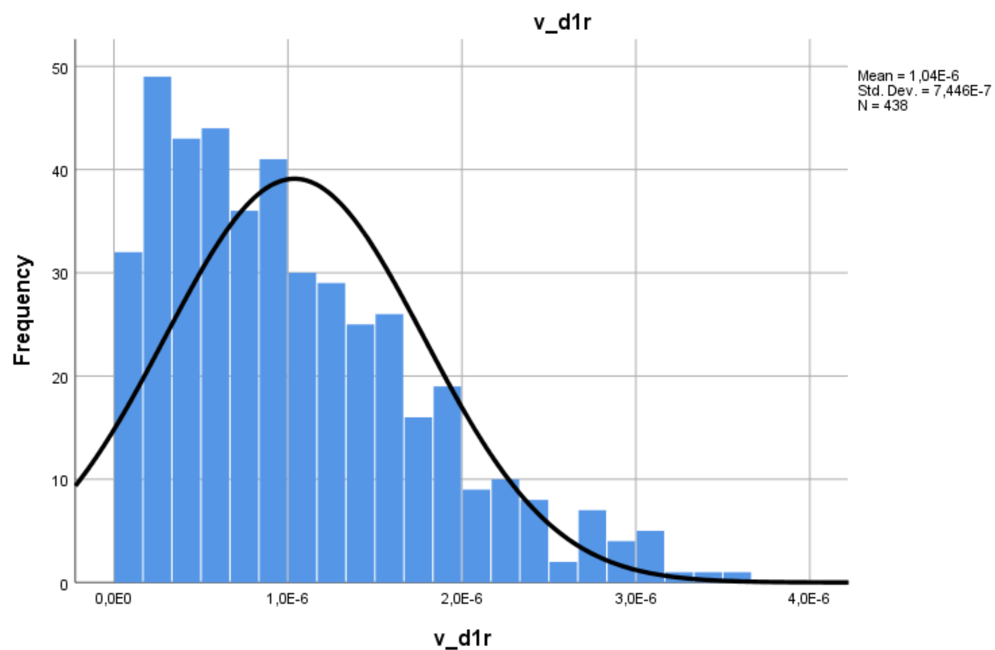

Figure 18. Frequencies of v_d1r

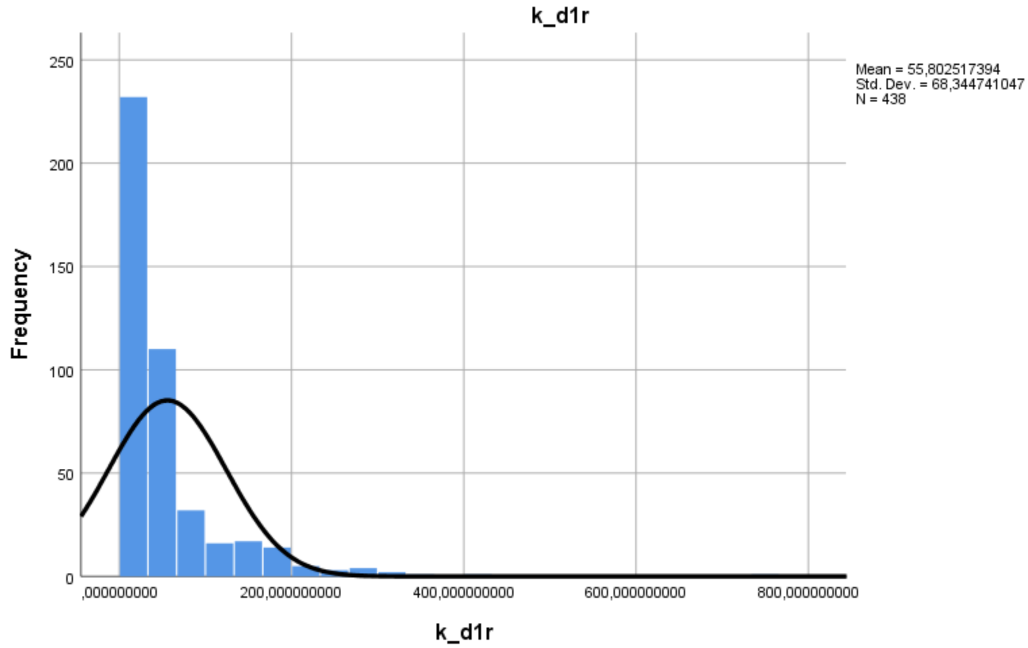

Figure 19. Frequencies of k_d1r

For the independent and dependent variables, we have that: most variables have similar behavior, unlike the time variable that has a high value of variance and kurtosis. Most variables have a low variance and kurtosis value. Regarding the asymmetry, the two variables with negative symmetry are the performance values (performanceN equals the mapped score and performanceP to the percentage).

Table 16. Descriptive statistics from Variables

		session ID	level	memory Span	performance N	performance P	time	gender C	age	severity
N	Valid	438	438	438	438	438	438	438	438	438
	Missing	0	0	0	0	0	0	0	0	0
Mean		4.77	2.03	4.06	2.62	.7937	147.20	.6689	11.43	.8539
Std. Deviation		2.980	.796	1.798	.608	.2320	151.06	.4711	2.36	.9034
Variance		8.878	.633	3.232	.370	.054	22819.44	.222	5.582	.816
Skewness		.508	.276	.177	-1.344	-1.083	2.344	-.721	.475	.292
Std. Error Skew.		.117	.117	.117	.117	.117	.117	.117	.117	.117
Kurtosis		-.754	-.633	-.359	.719	.779	5.990	-1.488	-1.462	-1.715
Std. Error Kurt.		.233	.233	.233	.233	.233	.233	.233	.233	.233
Minimum		1	1	0	1	.00	23	.00	9	.00
Maximum		12	4	9	3	1.00	973	1.00	15	2.00

The minimum and maximum values of each variable give us an account of their rating scale (Table 16). For example, the session varies between 1 and 12, while the level only handles 4 values. The variable time is the one having the greatest range of values from 23 seconds to 973 seconds. For the gender, the value Male corresponds to one and Female to zero. The age of the participants ranges between 9 and 15 years. Finally, in the severity of the diagnosis, we satisfy all the values of the variable according to the selection process of the children for this experiment.

- **Inferential Analysis**

From the descriptive statistics, variance in the data of the independent variables was, in most cases, zero. So, we could assume a normal distribution of data around the values of the mean. Still, we consider important to validate this conclusion by means of a test of normality.

For this test, we assume as a null hypothesis (H_0) that the distribution of the data is significantly normal with a significance greater than 0.05. According to the data in the Table 17, both normality tests (Kolmogorov-Smirnov[8] and Shapiro-Wilk[9]) have a significance of 0.0. Therefore, we reject the null hypothesis and declare, as the distribution of the normal curve is significantly different in the dependent variables, the distribution of the data is non-normal.

Table 17. Test of Normality for Dependent Variables

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
m_d1r	.156	438	.000	.732	438	.000
v_d1r	.093	438	.000	.932	438	.000
k_d1r	.249	438	.000	.602	438	.000
m_d1d1r	.080	438	.000	.941	438	.000
v_d1d1r	.090	438	.000	.932	438	.000
k_d1d1r	.252	438	.000	.604	438	.000

Once we have considered data with a non-normal distribution, we begin with the hypothesis testing of the relationships between the independent variables associated with the cognition and the dependent variable (EDA). For this purpose, we select the non-parametric tests, because they do not assume the same regarding the distribution of the parameters, they are more general than the parametric ones [140].

We validated a total of eight hypotheses: five for the principal independent variables and three for the secondary independent variables, with the aim of having more

[8] https://www.ibm.com/support/knowledgecenter/es/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/idh_ntk1.html
[9] https://www.ibm.com/support/knowledgecenter/es/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/idh_exam_plo.html

information when making decisions regarding the behavior of the data. The null hypotheses are the same: we declare that there is no correlation between the cognitive and/or demographic aspect of the child with the EDA signal (its tonic component - SCL) (Table 18).

Table 18. Description of hypotheses tested

Test	Null hypothesis	Alternative hypothesis
a. Session	H_{0a}: No correlation between the number of sessions and EDA.	H_{1a}: Correlation between the number of sessions and EDA.
b. Level (cognitive load)	H_{0b}: No correlation between the difficulty levels and EDA.	H_{1b}: Correlation between the difficulty levels and EDA.
c. Score (memory span)	H_{0c}: No correlation between memory span and EDA.	H_{1c}: Correlation between the memory span and EDA.
d. Performance	H_{0d}: No correlation between the performance (percentage) and EDA.	H_{1d}: Correlation between the performance (percentage) and EDA.
e. Time	H_{0e}: No correlation between the duration time and EDA.	H_{1e}: Correlation between the duration time and EDA.
f. Gender	H_{0f}: No differences between gender and EDA.	H_{1f}: There are differences between gender and EDA.
g. Age	H_{0g}: No differences between age and EDA.	H_{1g}: There are differences between age and EDA.
h. Severity (of diagnoses)	H_{0h}: No differences between the severity and EDA.	H_{1h}: There are differences between the severity and EDA.

For the five principal variables, the hypothesis test is a bivariate correlation analysis with the Kendall and Spearman coefficients (non-parametric), between the attributes that represent the EDA and each defined cognitive aspect. On the other hand, for the secondary variables, by their nature, we performed a non-parametric test of independent samples.

Table 19, Table 20, Table 21, Table 22, and Table 23 show the correlation values of the independent variable with each of the representative attributes of the signal.

Table 19. Correlations between Session and EDA

Correlations									
			m_d1r	v_d1r	k_d1r	m_d1d1r	v_d1d1r	k_d1d1r	session_id
Kendall's tau_b	session_id	Correlation coefficient	-.027	-.308**	.308**	.307**	-.308**	.308**	1.000
		Sig. (2-talied)	.426	.000	.000	.000	.000	.000	
		N	438	438	438	438	438	438	438

Spearman's rho	session_id	Correlation coefficient	-.041	-.432**	.434**	.433**	-.432**	.434**	1.000
		Sig. (2-talied)	.397	.000	.000	.000	.000	.000	
		N	438	438	438	438	438	438	438

*. Correlation is significant at the 0.05 level
**. Correlation is significant at the 0.01 level

Table 20. Correlations between Level and EDA

Correlations

			m_d1r	v_d1r	k_d1r	m_d1d1r	v_d1d1r	k_d1d1r	level
Kendall's tau_b	level	Correlation coefficient	-.042	-.314**	.304**	.307**	-.314**	.303**	1.000
		Sig. (2-talied)	.254	.000	.000	.000	.000	.000	
		N	438	438	438	438	438	438	438
Spearman's rho	level	Correlation coefficient	-.056	-.403**	.389**	.395**	-.402**	.389**	1.000
		Sig. (2-talied)	.239	.000	.000	.000	.000	.000	
		N	438	438	438	438	438	438	438

*. Correlation is significant at the 0.05 level
**. Correlation is significant at the 0.01 level

Table 21. Correlations between Memory span and EDA

Correlations

			m_d1r	v_d1r	k_d1r	m_d1d1r	v_d1d1r	k_d1d1r	Memory Span
Kendall's tau_b	memory Span	Correlation coefficient	.006	.068*	-.065	-.059	.069*	-.066	1.000
		Sig. (2-talied)	.870	.049	.059	.089	.044	.056	
		N	438	438	438	438	438	438	438
Spearman's rho	memory Span	Correlation coefficient	.009	.098*	-.094	-.085	.100*	-.094*	1.000
		Sig. (2-talied)	.854	.041	.050	.075	.037	.049	
		N	438	438	438	438	438	438	438

*. Correlation is significant at the 0.05 level
**. Correlation is significant at the 0.01 level

Table 22. Correlations between Performance and EDA

Correlations

			m_d1r	v_d1r	k_d1r	m_d1d1r	v_d1d1r	k_d1d1r	performance P
Kendall's tau_b	performance P	Correlation coefficient	-.013	.080*	-.076*	-.075*	.080*	-.075*	1.000
		Sig. (2-talied)	.701	.022	.030	.031	.022	.031	
		N	438	438	438	438	438	438	438
Spearman's rho	performance P	Correlation coefficient	-.017	.109*	-.104*	-.103*	.110*	-.103*	1.000
		Sig. (2-talied)	.721	.022	.030	.031	.022	.030	
		N	438	438	438	438	438	438	438

*. Correlation is significant at the 0.05 level
**. Correlation is significant at the 0.01 level

Table 23. Correlations between Time and EDA

Correlations

			m_d1r	v_d1r	k_d1r	m_d1d1r	v_d1d1r	k_d1d1r	Time
Kendall's tau_b	time	Correlation coefficient	.023	-.871**	.937**	.871**	-.872**	.940**	1.000

		Sig. (2-talied)	.472	.000	.000	.000	.000	.000	
		N	438	438	438	438	438	438	438
Spearman's rho	**time**	**Correlation coefficient**	.026	-.972**	.990**	.969**	-.972**	.991**	1.000
		Sig. (2-talied)	.584	.000	.000	.000	.000	.000	
		N	438	438	438	438	438	438	438

***. Correlation is significant at the 0.05 level**
****. Correlation is significant at the 0.01 level**

We summarize the results obtained in the hypothesis testing of the main variables in four groups, based on the interpretation of the correlation coefficients [141, p. 305]:

1. <u>Variables without correlation:</u> the variable Memory Span presented the lowest values of the correlation coefficient. For the Kendall coefficient, there is no correlation; for the Spearman coefficient, the correlation can be considered as very weakly negative, despite the coefficients are close to 0.1, and some significance was found.
2. <u>Variables with a week negative correlation:</u> for the Kendall coefficient, there is no correlation; for the Spearman coefficient, the correlation for the Performance variable (mean, variance and standard deviation of the first derivate) can be considered as very weakly negative, since the coefficients are a little bigger than 0.1 and significant at 0.05 level.
3. <u>Variables with a moderate positive correlation:</u> those variables where most of their attributes have a correlation coefficient between 0.25 and 0.5. The Session and Level variables are in this group, where the correlation is significant at the 0.01 level.
4. <u>Variables with very strong positive correlation:</u> the only variable with a correlation coefficient greater than 0.85 was Time.

Following the hypothesis testing for the secondary variables, the results show in Table 24,
Table 25, and
Table 26 divide the variables into two groups:

1. <u>Variables with significant differences:</u> for the variables Gender and Age we have significant differences (level of significance of 0.05) between each value presented by these variables and the EDA signal.

2. Variables without significant differences: for the Severity variable, there is no significant difference between the values this variable takes with respect to the EDA signal.

Furthermore, regarding the results obtained with the correlation analysis of the main variables, we decided to perform the independent t-tests on the Level and Performance variables to validate the differences of the values of these variables with respect to the EDA signal. These variables were qualified with positive moderate correlation and weak negative correlation respectively.

Table 27 and

Table 28 show that the variables Level and Performance have significant differences between the values of their respective ranges.

Table 24. Independent T-test with Gender

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of m_d1r is the same across categories of genderC.	Independent-Samples Mann-Whitney U Test	.796	Retain the null hypothesis
2	The distribution of v_d1r is the same across categories of genderC.	Independent-Samples Mann-Whitney U Test	.010	Reject the null hypothesis
3	The distribution of k_d1r is the same across categories of genderC.	Independent-Samples Mann-Whitney U Test	.019	Reject the null hypothesis
4	The distribution of m_d1d1r is the same across categories of genderC.	Independent-Samples Mann-Whitney U Test	.021	Reject the null hypothesis
5	The distribution of v_d1d1r is the same across categories of genderC.	Independent-Samples Mann-Whitney U Test	.009	Reject the null hypothesis
6	The distribution of k_d1d1r is the same across categories of genderC.	Independent-Samples Mann-Whitney U Test	.019	Reject the null hypothesis
Asymptotic significances are displayed. The significance level is .05				

Table 25. Independent T-test with Age

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of m_d1r is the same across categories of age.	Independent-Samples Kruskal-Wallis Test	.002	Reject the null hypothesis
2	The distribution of v_d1r is the same across categories of age.	Independent-Samples Kruskal-Wallis Test	.001	Reject the null hypothesis

3	The distribution of k_d1r is the same across categories of age.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis
4	The distribution of m_d1d1r is the same across categories of age.	Independent-Samples Kruskal-Wallis Test	.001	Reject the null hypothesis
5	The distribution of v_d1d1r is the same across categories of age.	Independent-Samples Kruskal-Wallis Test	.001	Reject the null hypothesis
6	The distribution of k_d1d1r is the same across categories of age.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis

Asymptotic significances are displayed. The significance level is .05

Table 26. Independent T-test with Severity

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of m_d1r is the same across categories of severity.	Independent-Samples Kruskal-Wallis Test	.572	Retain the null hypothesis
2	The distribution of v_d1r is the same across categories of severity.	Independent-Samples Kruskal-Wallis Test	.797	Retain the null hypothesis
3	The distribution of k_d1r is the same across categories of severity.	Independent-Samples Kruskal-Wallis Test	.839	Retain the null hypothesis
4	The distribution of m_d1d1r is the same across categories of severity.	Independent-Samples Kruskal-Wallis Test	.765	Retain the null hypothesis
5	The distribution of v_d1d1r is the same across categories of severity.	Independent-Samples Kruskal-Wallis Test	.796	Retain the null hypothesis
6	The distribution of k_d1d1r is the same across categories of severity.	Independent-Samples Kruskal-Wallis Test	.843	Retain the null hypothesis

Asymptotic significances are displayed. The significance level is .05

Table 27. Independent T-test (retesting) with Level

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of m_d1r is the same across categories of level.	Independent-Samples Kruskal-Wallis Test	.431	Retain the null hypothesis
2	The distribution of v_d1r is the same across categories of level.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis
3	The distribution of k_d1r is the same across categories of level.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis
4	The distribution of m_d1d1r is the same across categories of level.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis

5	The distribution of v_d1d1r is the same across categories of level.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis
6	The distribution of k_d1d1r is the same across categories of level.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis
Asymptotic significances are displayed. The significance level is .05				

Table 28. Independent T-test (retesting) with Performance

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of m_d1r is the same across categories of performanceN.	Independent-Samples Kruskal-Wallis Test	.309	Retain the null hypothesis
2	The distribution of v_d1r is the same across categories of performanceN.	Independent-Samples Kruskal-Wallis Test	.019	Reject the null hypothesis
3	The distribution of k_d1r is the same across categories of performanceN.	Independent-Samples Kruskal-Wallis Test	.038	Reject the null hypothesis
4	The distribution of m_d1d1r is the same across categories of performanceN.	Independent-Samples Kruskal-Wallis Test	.029	Reject the null hypothesis
5	The distribution of v_d1d1r is the same across categories of performanceN.	Independent-Samples Kruskal-Wallis Test	.019	Reject the null hypothesis
6	The distribution of k_d1d1r is the same across categories of performanceN.	Independent-Samples Kruskal-Wallis Test	.039	Reject the null hypothesis
Asymptotic significances are displayed. The significance level is .05				

To conclude with the statistical analysis, in Table 29 we present a summary of the hypotheses retained and rejected according to the results of each bivariate analysis and independent samples tests.

Table 29. Hypothesis retained and rejected

Test	Probed hypothesis
a. Session	H_{1a}: Moderate positive correlation (Reject null hypothesis)
b. Level	H_{1b}: Moderate positive correlation (Reject null hypothesis)
c. Score (memory span)	H_{0c}: No correlation (Retain null hypothesis)
d. Performance	H_{0d}: Weak negative correlation (Reject null hypothesis)

e. Time	H_{1e}: Very strong positive correlation (Reject null hypothesis)
f. Gender	H_{1f}: Significant difference (Reject null hypothesis)
g. Age	H_{1g}: Significant difference (Reject null hypothesis)
h. Severity (of diagnoses)	H_{0h}: No significant difference (Retain null hypothesis)
i. Level (retested)	H_{0i}: Significant difference (Reject null hypothesis)
j. Performance (retested)	H_{0j}: Significant difference (Reject null hypothesis)

5.2.5. Machine Learning

To continue with the third step of the signal processing (feature selection), starting with the choice of the most appropriate method for this purpose, we detail the second approach from the paradigm of machine learning. Machine learning is an application of artificial intelligence (AI) for providing the systems with the ability to automatically learn and improve from experience without being explicitly programmed. The aim consists in allowing computers to learn automatically without human intervention [142].

For this experiment, we use some supervised machine learning algorithms that apply what they learned in the past to new data using labeled samples to predict future events. These algorithms are grouped into regression and classification, according to the output variable used to label the data [143]. Since our labeling variables are qualitative, we choose to work with classification algorithms.

Unlike statistical analysis and for our particular case, machine learning algorithms receive the features of the signal to try to find patterns in the data that allow us to classify correctly according to the given labels, that is, the information about of the cognition we know (Figure 20).

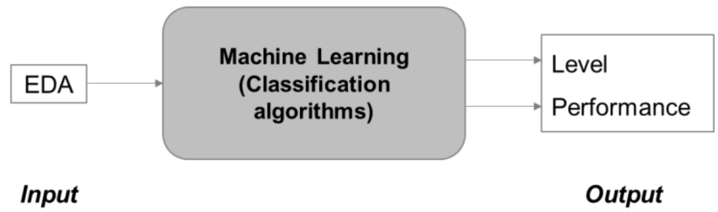

Figure 20. Input and output variables for Machine Learning

To perform the analysis from machine learning algorithms, we define that the two labels to analyze are the Level and Performance. We make the decision based on:

- The range of values of the variables is small, which would avoid an under sampling of the instances of the dataset.
- They are representative variables of the set of aspects of cognition.

We base our selection criteria of the classification algorithms used for this experiment in the extensive use of these in the literature. We performed the analysis with nine different algorithms and an optimization function:

- Ensemble meta-algorithms:
  - Boosting [144]
  - Bootstrap Aggregating – Bagging [144]
  - Random Forest [145]
- Support Vector Machines algorithms:
  - Sequential Minimal Optimization – SMO [146]
  - LibSVM [146]
  - GridSearch (Optimization function) [147]
- Neural Networks algorithms:
  - Multilayer Perceptron [148]
- Decision Trees algorithms:
  - J48 [149]
- Probabilistic algorithms:
  - NaiveBayes [150]
- Non-parametric algorithms:
  - K Nearest Neighbors – KNN (iBk) [151]

In a preliminary way and following the protocol found in most of the literature [118], [126], we obtain a set of the results from the classification models generated by each algorithm for our dataset. It is important to remember that the characteristics of our dataset include 436 instances labeled according to level and performance variables, along with a total of 360 attributes.

Having this first version of the dataset, the best algorithm for the performance classification was LibSVM with the GridSearch optimization function with a correctly classified percentage of instances of 68.26% (299 instances). In the classification of the Level, the best algorithm was SMO with 47.47% of correctly classified instances (209 instances).

Nevertheless, the results of this preliminary tests are not very reliable due to some data quality issues according to Corrales et al. [152]:

- Imbalanced class: is considered when a dataset exhibits an unequal distribution between its classes. An imbalanced dataset contributes to decrease the accuracy and the quality of the learning.
- High dimensionality: in presence of a large number of attributes, a classification learning model tends to over fit the model, resulting in lowers performance.
- Amount of data: the total available data for training a classifier and it is highly related with high dimensionality. Small datasets can build inaccurate models.

We identify that our dataset suffers from these three issues with its data quality. Table 30 and Figure 21 show the preliminary distribution of the dataset with the Level class. An important clarification: although the levels of difficulty in HapHop-Physio are five, the levels that we reported in the dataset are only four, given that within the time of the study with the children, none of them reached the fifth level of difficulty. Table 31 and Figure 22 show the preliminary distribution of the dataset with the Performance class.

Table 30. Distribution of the Level class

N°	Label	Count	Weight
1	1	120	120.0
2	2	197	197.0
3	3	109	109.0
4	4	12	12.0

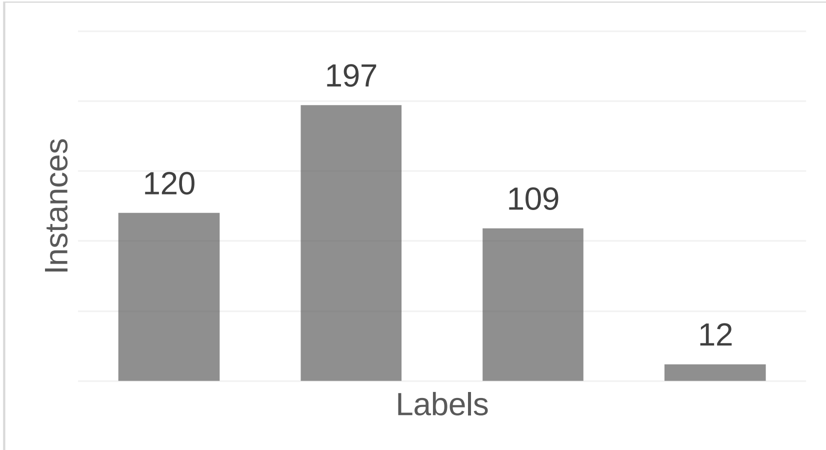

Figure 21. Distribution of the Level class

Table 31. Distribution of the Performance class

N°	Label	Count	Weight
1	1	120	120.0
2	2	197	197.0
3	3	109	109.0
4	4	12	12.0

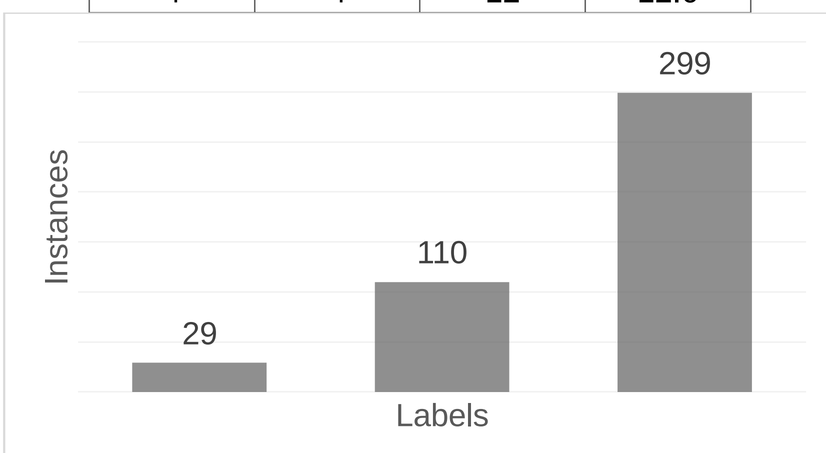

Figure 22. Distribution of the Performance class

Regarding the first problem, we balanced the classes by generating synthetic instances through oversampling. For the second problem, we use an approximation of the selection of the relevant attributes for the classifier by means of a Wrapping approach [152].

For the dataset with the Level label, class balancing increased the number of instances to 731, while the reduction of the dimensionality by the Wrapper approach (with the RandomForest algorithm) decreased the number of attributes to 17. We should note that, although HapHop-Physio has five levels of difficulty, the dataset generated by the interaction of children with the game, only has four levels or four labeling options of the Level class; this is because no child managed to complete beyond level 4 in the game. For the case of the dataset with the Performance class, the number of instances increased to 864 and of attributes decreased to 28 (also with the RandomForest algorithm).

- **Final Classification Results**

After solving the issues, we run the classification algorithms for the second version of the dataset. In Table 32 and

Table 33, we observed the final results of the classification models generated by the algorithms. Both for the Level label and for the Performance label, the best algorithm was RandomForest, although the best precision percentage was the dataset with the Performance class. In Figure 23 and Figure 24 we show the results buffer from the classification model performed in Weka[10]. It is important to highlight from these results: the instances metrics, the confusion matrix and the kappa statistics.

Table 32. Classification results with Level class

Algorithm	Correctly Classified Instances	Percentage	Incorrectly Classified Instances	Percentage
AdaBoostM1 (J48)	485	66,3475%	246	33,6525%
Bagging (J48)	478	65,3899%	253	34,6101%
RandomForest	535	73,1874%	196	26,8126%
SMO	385	52,6676%	346	47,3324%
LibSVM (Default)	397	54,3092%	334	45,6908%
GridSearch (LibSVM)	197	26,9494%	534	73,0506%
Multilayer Perceptron	428	58,5499%	303	41,4501%
J48 (Default)	449	61,4227%	282	38,5773%
NaiveBayes (Default)	352	48,1532%	379	51,8468%
IBk	498	68,1259%	233	31,8741%

[10] https://www.cs.waikato.ac.nz/ml/weka/

Table 33. Classification results with Performance class

Algorithm	Correctly Classified Instances	Percentage	Incorrectly Classified Instances	Percentage
AdaBoostM1 (J48)	617	71,4120%	247	28,5880%
Bagging (J48)	625	72,3380%	239	27,6620%
RandomForest	682	78,9352%	182	21,0648%
SMO	415	48,0324%	449	51,9676%
LibSVM (Default)	448	51,8519%	416	48,1481%
GridSearch (LibSVM)	299	34,6065%	565	65,3935%
Multilayer Perceptron	491	56,8287%	373	43,1713%
J48 (Default)	527	60,9954%	337	39,0046%
NaiveBayes (Default)	434	50,2315%	430	49,7685%
iBk	566	65,5093%	298	34,4907%

To guarantee the quality in the classification accuracy of an algorithm, there are different metrics to establish it. One of those metrics is the Kappa statistic. In essence, this metric is a measure of how closely the instances classified by the machine learning algorithm matched the data labeled as ground truth, controlling for the accuracy of a random classifier as measured by the expected accuracy [153].

```
RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.41 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         535               73.1874 %
Incorrectly Classified Instances       196               26.8126 %
Kappa statistic                          0.6428
Mean absolute error                      0.2139
Root mean squared error                  0.3123
Relative absolute error                 57.0716 %
Root relative squared error             72.1434 %
Total Number of Instances              731

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,683    0,098    0,695      0,683    0,689      0,589    0,882     0,730     1
                 0,538    0,109    0,646      0,538    0,587      0,457    0,831     0,594     2
                 0,736    0,117    0,663      0,736    0,698      0,598    0,894     0,745     3
                 0,989    0,034    0,904      0,989    0,944      0,927    0,998     0,995     4
Weighted Avg.    0,732    0,090    0,726      0,732    0,726      0,639    0,900     0,762

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 123  28  24   5 |   a = 1
  40 106  41  10 |   b = 2
  14  28 128   4 |   c = 3
   0   2   0 178 |   d = 4
```

Figure 23. Results buffer from RandomForest with Level

The Kappa statistic for the RandomForest classifier in the Level data set is 0.642 (Figure 23), while for the Performance data set it has a value of 0.684 (Figure 24). According to [154], these values are classified within the range of substantial agreement. Therefore, we can conclude that the accuracy of the classifier is fairly good.

```
RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.49 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         682               78.9352 %
Incorrectly Classified Instances       182               21.0648 %
Kappa statistic                          0.684
Mean absolute error                      0.2925
Root mean squared error                  0.3498
Relative absolute error                 65.8422 %
Root relative squared error             74.2168 %
Total Number of Instances              864

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,934    0,078    0,858      0,934    0,894      0,839    0,973     0,943     Bajo
                 0,720    0,134    0,715      0,720    0,717      0,585    0,883     0,761     Medio
                 0,712    0,103    0,786      0,712    0,747      0,625    0,879     0,782     Alto
Weighted Avg.    0,789    0,105    0,787      0,789    0,787      0,684    0,912     0,829

=== Confusion Matrix ===

   a   b   c   <-- classified as
 271  14   5 |   a = Bajo
  24 198  53 |   b = Medio
  21  65 213 |   c = Alto
```

Figure 24. Result buffer with RandomForest for Performance

5.2.6. Comments to the Results

According to the steps considered for the signal processing, we found more relevant the analysis of the signal from the machine learning approach. With the tasks we carried out within the analysis of the classifiers, we addressed the last two steps of the processing: the attribute selection and the final extraction of the relevant information by means of classification.

Despite the correlation values obtained during the statistical analysis were not decisive to perform the feature selection step, it allowed us to rise new hypothesis, such as what are the relevant features when analyzing EDA considering that this is a non-stationary signal (variable over time). Using the machine learning approach allows us to apply a more robust statistical analysis, considering many features that are possible applying for instance the Wavelet transformations. In consequence, the classification by means of machine learning algorithms gave encouraging results for the classification of the variable Performance (one aspect of cognition).

Thus, with the final results from our experiment, reported in this section, we obtain one last conclusion of this master's work: we were able to find that there is a relationship between cognition and the EDA signal when we frame the process in the objective method.

Summary

In this chapter, we describe in detail two evaluation resources for the validation of this research work. In the first place, a validation of academic type towards the first contribution delivered in this master's thesis: our conceptual framework and objective method for the evaluation of the UX.

In the second place, and following an experimental design according to the steps defined for the signal processing, we validated the proposed procedure, which aimed to find the relationship between cognition and the physiological signal EDA.

In this experimental design, we approached two sets of estimates for the analysis of the data: a statistical analysis (descriptive and inferential) in which we observed and understood the behavior of the data of the dataset generated from the segmented signals, and an analysis from supervised machine learning in order to find patterns of relationship between EDA signal data and aspects of cognition.

The final results of the complete validation process demonstrate the fulfillment of the objectives proposed in the formulation of this research project. The next and final chapter poses the conclusions and future works derivate from the results obtain in this master’s thesis.

Chapter 6

Conclusions and Future Works

In this dissertation, we have presented a method for the objective evaluation of UX from the conceptual framework built to identify the measurable aspects of UX. Our method links the physiological signals measuring the internal state of the user with the psychological events defined from the psychophysiology.

In this chapter, we present the conclusions obtained from the development of this master's thesis, especially considering the results of the evaluation process. Furthermore, we propose some possible future works that could improve the accuracy of our procedure to relate the EDA data with the cognition.

6.1. Conclusions

Based on the results of this research, and considering the analysis of our method, we can conclude that:

- Carrying out a systematic mapping allowed us to recognize the failures that were occurring in the methods reported around the evaluation of the UX. There was no clear differentiation between the evaluation of the user and the evaluation of the system and, therefore, there was confusion about the evaluative aspects of the UX corresponding to each of the subjects involved in the interaction.

- The systematic literature review provided us with a compendium of the type of applications that, like HapHop-Physio, help in the rehabilitation processes of children with different needs in their physical, mental and emotional health. We understood that the concept of UX plays a vital role in the design and evaluation of these applications, therefore, we start from the need to frame our project in a comprehensive conceptual framework, despite it was not planned in the project statement at the beginning of the Master work.

- Our conceptual framework organized the most relevant concepts within the UX field, taking as reference several standards. The framework defined in a more clearly and precisely way, the differences in the UX evaluation seen from the user and the system perspectives. The conceptual framework:
    - Relates the UX evaluation methods from the user's perspective.
    - Relates the standards to determine the quality of the system from the user's point of view.

- The objective method obtained for the UX evaluation was thought based on the ISO quality standards for evaluating products, systems or services in software engineering, which makes it a replicable method in other contexts. The method also provides clarity regarding the procedure followed to evaluate the UX in a given context.

- The formulation of the conceptual framework and the objective method will allow future research in the objective area of the UX to have a broad frame of reference for the selection of the measurable aspects of UX according to the purpose of interaction, the design of the system and the needs of the user.

- The execution flow in the tasks of the obtained procedure agrees with the information flow of the objective method. Besides, it is aligned with the design objectives (cognitive rehabilitation game) of HapHop-Physio.

- Obtaining the dataset is one of the main contributions of this research work. Although it is a small dataset, some of its advantages include:
    - Data collected in a real environment with conventional users of the therapies and potential users of the system.
    - No part of the data preparation process, such as segmentation and labeling, involved subjective methods that could increase the risk of information bias.
    - The data quality was continually verified in the preparation of the data, through the monitoring of previously established rules in the literature, such as outliers and missing values.

- In the machine learning analysis, we obtained an acceptable classification percentage. This percentage is due to multiple causes, one of the most important is the size of the dataset. However, this percentage is a good indicator of the relationships and/or patterns that machine learning can find to relate the aspects associated with cognition from the EDA signal. A correlation between EDA and cognition aspects was also demonstrated in the statistical analysis.

- According to the state of the art, no other work has approached the study of the relationship between cognition and EDA to determine UX. Moreover, no one have related this physiological signal with the metrics obtained by any given system, which are used for the definition of some aspects of cognition.

- Finally, the proposed method and the instantiated procedure allowed the verification of the relationship hypothesis between the EDA and cognition in a reliable way (classification performance and statistical correlational analyisis), so we can assure that we have managed to determine the UX on the user in a completely objective way, thus fulfilling the stipulated in the objectives of this master's thesis.

6.2. Future works

The validation process showed various possibilities for improving the performance and usefulness of our procedure. Considering the results, we obtained in this dissertation, we propose as future works:

- Increase the size of the dataset to improve the accuracy of the classification model obtained. In the framework of the HapHop-Physio project (ID 4441), we will carry out more pilot tests with the game. From these exercises, we will obtain more data, both physiological signals and other aspects of cognition as attention.

- Make comparisons of personal and universal models for the recognition of individual and collective patterns in the physiological signal EDA, in order to relate this data to other aspects of cognition.

- Evaluate the EDA classification model for the performance aspect of cognition in real activities of daily life of children, such as performance in the school day or checking the rehabilitation activities at home that are part of the treatments for children with SLD.

- Instantiate the procedure towards the recognition of complex emotions in children with SLD, such as frustration, from the set of signals obtained.

- Instantiate the procedure towards the recognition of different behaviors in children with SLD from the set of signals obtained.

- Nourish the dataset with the other physiological signals obtained with the E4 device, such as temperature and heart rate.

References

[1] J. Kort, A. Vermeeren, and J. E. Fokker, “Conceptualizing and measuring user experience,” *UX Manif.*, p. 57, 2007.

[2] A. Allam and H. M. Dahlan, “User experience: Challenges and opportunities,” *J. Inf. Syst. Res. Innov.*, vol. 3, pp. 28–36, 2013.

[3] “All About UX.” [Online]. Available: http://www.allaboutux.org/. [Accessed: 31-Oct-2016].

[4] S. Rajeshkumar, R. Omar, and M. Mahmud, “Taxonomies of User Experience (UX) evaluation methods,” in *Research and Innovation in Information Systems (ICRIIS), 2013 International Conference on*, 2013, pp. 533–538.

[5] M. Zarour and M. Alharbi, “User experience framework that combines aspects, dimensions, and measurement methods,” *Cogent Eng.*, no. just-accepted, p. 1421006, 2017.

[6] V. Roto, A. Vermeeren, K. Väänänen-Vainio-Mattila, and E. Law, “User Experience Evaluation – Which Method to Choose?,” in *Human-Computer Interaction – INTERACT 2011*, P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, Eds. Springer Berlin Heidelberg, 2011, pp. 714–715.

[7] P. Kashfi, R. Feldt, A. Nilsson, and R. B. Svensson, “A Conceptual UX-Aware Model of Requirements,” in *Human-Centered and Error-Resilient Systems Development*, Springer, Cham, 2016, pp. 234–245.

[8] C. Rico-Olarte, D. López, and S. Kepplinger, “User Experience Evaluation Methods in Games for Children with Cognitive Disabilities: A Systematic Review,” in *Proc. of 1st International Young Researcher Summit on Quality of Experience in Emerging Multimedia Services (QEEMS 2017)*, Erfurt, Germany, 2017.

[9] C. Rico-Olarte, S. Narváez, C. Farinango, D. M. López, and P. S. Pharow, “HapHop-Physio A Computer Game to Support Cognitive Therapies in Children,” *Psychol. Res. Behav. Manag.*, vol. 10, pp. 1–9, 2017.

[10] M. Rosselli, E. Matute, and A. Ardila, *Neuropsicología del desarrollo infantil*. Editorial El Manual Moderno, 2010.

[11] M. Rosselli-Cock *et al.*, “Evaluación Neuropsicológica Infantil (ENI): una batería para la evaluación de niños entre 5 y 16 años de edad. Estudio normativo colombiano,” *Rev. Neurol.*, vol. 38, no. 8, pp. 720–731, 2004.

[12] G. Blázquez Almería *et al.*, “Resultados del cribado de la sintomatología del trastorno por déficit de atención con o sin hiperactividad en el ámbito escolar mediante la escala EDAH,” *Rev Neurol*, vol. 41, no. 10, pp. 586–90, 2005.

[13] A. Urzúa, M. Ramos, C. Alday, and A. Alquinta, “Madurez neuropsicológica en preescolares: propiedades psicométricas del test CUMANIN,” *Ter. Psicológica*, vol. 28, no. 1, pp. 13–25, 2010.

[14] C. Hannaford, *Aprender moviendo el cuerpo*. Editorial Pax México, 2009.

[15] D. E. Sarver, M. D. Rapport, M. J. Kofler, J. S. Raiker, and L. M. Friedman, “Hyperactivity in Attention-Deficit/Hyperactivity Disorder (ADHD): Impairing Deficit or Compensatory Behavior?,” *J. Abnorm. Child Psychol.*, vol. 43, no. 7, pp. 1219–1232, Oct. 2015.

[16] Y. Jabareen, “Building a conceptual framework: philosophy, definitions, and procedure,” *Int. J. Qual. Methods*, vol. 8, no. 4, pp. 49–62, 2009.

[17] M. Hassenzahl, “The hedonic/pragmatic model of user experience,” *UX Manif.*, vol. 10, pp. 10–14, 2007.

[18] Instituto Colombiano de Bienestar Familiar, “Discapacidad cognitiva.” 2008.

[19] T. Susi, M. Johannesson, and P. Backlund, *Serious games: An overview*. Institutionen för kommunikation och information, 2007.

[20] L. Caroux, K. Isbister, L. Le Bigot, and N. Vibert, “Player–video game interaction: A systematic review of current concepts,” *Comput. Hum. Behav.*, vol. 48, pp. 366–381, Jul. 2015.

[21] F. E. P. Corredor, J. E. M. Cardona, and O. A. H. Gallo, “Diseño de videojuegos serios para la salud,” *Rev. Páginas*, vol. 0, no. 95, pp. 127–141, 2014.

[22] L. Lopez-Samaniego and B. Garcia-Zapirain, “A Robot-Based Tool for Physical and Cognitive Rehabilitation of Elderly People Using Biofeedback,” *Int. J. Environ. Res. Public. Health*, vol. 13, no. 12, Nov. 2016.

[23] J. M. Burns, M. Webb, L. A. Durkin, and I. B. Hickie, “Reach Out Central: a serious game designed to engage young men to improve mental health and wellbeing,” *Med. J. Aust.*, vol. 192, no. 11 Suppl, pp. S27-30, Jun. 2010.

[24] M. Frutos-Pascual, B. G. Zapirain, and K. C. Buldian, “Adaptive Cognitive Rehabilitation Interventions Based on Serious Games for Children with ADHD Using Biofeedback Techniques: Assessment and Evaluation,” in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, ICST, Brussels, Belgium, Belgium, 2014, pp. 321–324.

[25] R. Rocha, L. P. Reis, P. A. Rego, P. M. Moreira, and B. M. Faria, “New forms of interaction in serious games for cognitive rehabilitation: Implementation and usability study,” in *2016 11th Iberian Conference on Information Systems and Technologies (CISTI)*, 2016, pp. 1–6.

[26] W. L. Johnson, H. Vilhjalmsson, and S. Marsella, “Serious Games for Language Learning: How Much Game, How Much AI?,” in *Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning Through Intelligent and Socially Informed Technology*, Amsterdam, The Netherlands, The Netherlands, 2005, pp. 306–313.

[27] R. L. Hazlett, “Measuring emotional valence during interactive experiences: boys at video game play,” in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 1023–1026.

[28] C. Rico-Olarte, D. López, B. Blobel, and S. Kepplinger, “User Experience Evaluations in Rehabilitation Video Games for Children: A Systematic Mapping of the Literature.,” *Stud. Health Technol. Inform.*, vol. 243, p. 13, 2017.

[29] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, “An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments,” in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–6.

[30] S. Gualeni, D. Janssen, and L. Calvi, “How Psychophysiology Can Aid the Design Process of Casual Games: A Tale of Stress, Facial Muscles, and Paper Beasts,” in *Proceedings of the International Conference on the Foundations of Digital Games*, New York, NY, USA, 2012, pp. 149–155.

[31] M. Moghim, R. Stone, P. Rotshtein, and N. Cooke, “Adaptive virtual environments: A physiological feedback HCI system concept,” in *2015 7th Computer Science and Electronic Engineering Conference (CEEC)*, 2015, pp. 123–128.

[32] Y. Bian *et al.*, “A Physiological Evaluation Model for Flow-Experience in VR Games: Construction and Preliminary Test,” in *2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI)*, 2015, pp. 244–249.

[33] A. S. Billis, E. I. Konstantinidis, A. I. Ladas, M. N. Tsolaki, C. Pappas, and P. D. Bamidis, “Evaluating affective usability experiences of an exergaming platform for seniors,” in *2011 10th International Workshop on Biomedical Engineering*, 2011, pp. 1–4.

[34] A. Alloni, D. Tost, S. Panzarasa, C. Zucchella, and S. Quaglini, “Applying 3D Graphics to Computerized Cognitive Rehabilitation,” in *ICTs for Improving Patients Rehabilitation Research Techniques*, 2014, pp. 115–128.

[35] D. Tost, A. von Barnekow, E. Felix, S. Pazzi, S. Puricelli, and S. Bottiroli, “Early Detection of Cognitive Impairments with the Smart Ageing Serious Game,” in *ICTs for Improving Patients Rehabilitation Research Techniques*, 2014, pp. 183–195.

[36] P. Hart, R. Bierwirth, G. Fulk, and E. Sazonov, “The design and evaluation of an activity monitoring user interface for people with stroke,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 5908–5911, 2014.

[37] G. Burdea *et al.*, “Feasibility study of the BrightBrainer™ integrative cognitive rehabilitation system for elderly with dementia,” *Disabil. Rehabil. Assist. Technol.*, vol. 10, no. 5, pp. 421–432, 2015.

[38] R. Navarro-Prieto and E. Parra, “Towards a Comprehensive Definition of User Experience for 3D Content,” in *User Centric Media*, 2010, pp. 51–59.

[39] A. R. Dores *et al.*, “Serious Games: Are They Part of the Solution in the Domain of Cognitive Rehabilitation?,” in *Serious Games Development and Applications*, 2011, pp. 95–105.

[40] C. S. Crawford, M. Andujar, F. Jackson, S. Remy, and J. E. Gilbert, “User experience evaluation towards cooperative brain-robot interaction,” in *International Conference on Human-Computer Interaction*, 2015, pp. 184–193.

[41] T. Y. Chuang and W.-F. Chen, “Effect of Computer-Based Video Games on Children: An Experimental Study,” in *2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL’07)*, 2007, pp. 114–118.

[42] T. Mondéjar, R. Hervás, J. M. Latorre, I. G. Diaz, and J. Bravo, “Analysis of EEG Frequency Bands During Typical Mechanics of Platform-Videogames,” in *Ubiquitous Computing and Ambient Intelligence*, 2016, pp. 306–317.

[43] R. Colpani and M. R. P. Homem, “An innovative augmented reality educational framework with gamification to assist the learning process of children with intellectual disabilities,” in *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 2015, pp. 1–6.

[44] A. Barco *et al.*, “A Robotic Therapy for Children with TBI,” in *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, Piscataway, NJ, USA, 2013, pp. 75–76.

[45] S. Ghisio *et al.*, “An open platform for full body interactive sonification exergames,” in *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*, 2015, pp. 168–175.

[46] E. Sürer, “Video-games based framework designed for the cognitive rehabilitation of children with Down Syndrome,” in *2016 24th Signal Processing and Communication Application Conference (SIU)*, 2016, pp. 1821–1824.

[47] V. Brezinka, “Computer games supporting cognitive behaviour therapy in children,” *Clin. Child Psychol. Psychiatry*, vol. 19, no. 1, pp. 100–110, Jan. 2014.

[48] S. Mäkelä, R. Bednarik, and M. Tukiainen, “Evaluating User Experience of Autistic Children Through Video Observation,” in *CHI ’13 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2013, pp. 463–468.

[49] D. Kamińska, T. Sapiński, and A. Pelikant, “Automatic behavioural therapy tool for children with Asperger’s syndrome - Paradigm,” in *2013 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2013, pp. 172–175.

[50] N. A. A. Zaki, T. S. M. T. Wook, and K. Ahmad, “Towards Developing a Therapeutic Serious Game Design Model for Stimulating Cognitive Abilities: A Case for Children with Speech and Language Delay,” in *Computational Intelligence in Information Systems*, 2016, pp. 259–269.

[51] M. J. Boivin, N. Nakasujja, A. Sikorskii, R. O. Opoka, and B. Giordani, “A Randomized Controlled Trial to Evaluate if Computerized Cognitive Rehabilitation Improves Neurocognition in Ugandan Children with HIV,” *AIDS Res. Hum. Retroviruses*, vol. 32, no. 8, pp. 743–755, Apr. 2016.

[52] S. Cano, D. M. Alghazzawi, J. M. Arteaga, H. M. Fardoun, C. A. Collazos, and V. B. Amador, “Applying the information search process model to analyze aspects in the design of serious games for children with hearing impairment,” *Univers. Access Inf. Soc.*, pp. 1–13, Jan. 2017.

[53] G. Sim, M. Nouwen, J. Vissers, M. Horton, K. Slegers, and B. Zaman, “Using the MemoLine to capture changes in user experience over time with children,” *Int. J. Child-Comput. Interact.*, vol. 8, pp. 1–14, May 2016.

[54] F. Martínez, C. Barraza, N. González, and J. González, "KAPEAN: Understanding Affective States of Children with ADHD," *Educ. Technol. Soc.*, vol. 19, no. 2, pp. 18–28, 2016.

[55] J. Marco, S. Baldassarri, and E. Cerezo, "NIKVision: Developing a Tangible Application for and with Children.," *J UCS*, vol. 19, no. 15, pp. 2266–2291, 2013.

[56] D. Z. Blandón, J. E. Muñoz, D. S. Lopez, and O. H. Gallo, "Influence of a BCI neurofeedback videogame in children with ADHD. Quantifying the brain activity through an EEG signal processing dedicated toolbox," in *2016 IEEE 11th Colombian Computing Conference (CCC)*, 2016, pp. 1–8.

[57] N. Ibrahim, W. Fatimah, W. Ahmad, and A. Shafie, "User Experience Study on Folktales Mobile Application for Children's Education," in *2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies*, 2015, pp. 353–358.

[58] C. S. González-González and V. Navarro-Adelantado, "Methods and Techniques for Evaluating the Emotional Experiences of Children with Active Videogames," in *Proceedings of the XVI International Conference on Human Computer Interaction*, New York, NY, USA, 2015, pp. 16:1–16:2.

[59] G. Sim and M. Horton, "Investigating Children's Opinions of Games: Fun Toolkit vs. This or That," in *Proceedings of the 11th International Conference on Interaction Design and Children*, New York, NY, USA, 2012, pp. 70–77.

[60] A. Landowska and J. Miler, "Limitations of emotion recognition in software user experience evaluation context," in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2016, pp. 1631–1640.

[61] S. Bampatzia, A. Antoniou, G. Lepouras, E. Roumelioti, and V. Bravou, "Comparing game input modalities: A study for the evaluation of player experience by measuring emotional state and game usability," in *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, 2015, pp. 530–531.

[62] C. Schroeter *et al.*, "Realization and user evaluation of a companion robot for people with mild cognitive impairments," in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 1153–1159.

[63] M. Joselli, F. Binder, E. Clua, and E. Soluri, "Mindninja: Concept, Development and Evaluation of a Mind Action Game Based on EEGs," in *2014 Brazilian Symposium on Computer Games and Digital Entertainment*, 2014, pp. 123–132.

[64] V. Brezinka, "Ricky and the Spider – a video game to support cognitive behavioural treatment of children with obsessive-compulsive disorder," *Clin. Neuropsychiatry*, vol. 10, no. 3, pp. 6–12, 2013.

[65] V. Roto, E. L.-C. Law, A. Vermeeren, and J. Hoonhout, "User Experience White Paper," Dagstuhl, Germany, 2011.

[66] I. Carrillo, V. Meza-Kubo, A. L. Morán, G. Galindo, and E. García-Canseco, "Processing EEG Signals Towards the Construction of a User Experience Assessment Method," in *Ambient Intelligence for Health*, 2015, pp. 281–292.

[67] T. Keskinen *et al.*, "Evaluating the experiential user experience of public display applications in the wild," in *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, 2013, p. 7.

[68] E. L.-C. Law, A. P. Vermeeren, M. Hassenzahl, and M. Blythe, "Towards a UX manifesto," in *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2*, 2007, pp. 205–206.

[69] E. Law, V. Roto, A. P. Vermeeren, J. Kort, and M. Hassenzahl, "Towards a shared definition of user experience," in *CHI'08 extended abstracts on Human factors in computing systems*, 2008, pp. 2395–2398.

[70] K. Väänänen-Vainio-Mattila, V. Roto, and M. Hassenzahl, "Now let's do it in practice: user experience evaluation methods in product development," in *CHI'08 extended abstracts on Human factors in computing systems*, 2008, pp. 3961–3964.

[71] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience: a survey approach," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 719–728.

[72] M. Hassenzahl, "User experience (UX): towards an experiential perspective on product quality," in *Proceedings of the 20th Conference on l'Interaction Homme-Machine*, 2008, pp. 11–15.

[73] M. Hassenzahl and N. Tractinsky, "User experience-a research agenda," *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 91–97, 2006.

[74] A. Wiklund-Engblom, M. Hassenzahl, A. Bengs, and S. Sperring, "What needs tell us about user experience," *Hum.-Comput. Interaction–INTERACT 2009*, pp. 666–669, 2009.

[75] M. Thüring and S. Mahlke, "Usability, aesthetics and emotions in human–technology interaction," *Int. J. Psychol.*, vol. 42, no. 4, pp. 253–264, 2007.

[76] R. L. Mandryk, "Physiological measures for game evaluation," *Game Usability Advice Experts Adv. Play. Exp.*, pp. 207–235, 2008.

[77] K. Boehner, R. DePaula, P. Dourish, and P. Sengers, "How emotion is made and measured," *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 4, pp. 275–291, 2007.

[78] D. Stone, C. Jarrett, M. Woodroffe, and S. Minocha, *User interface design and evaluation*. Morgan Kaufmann, 2005.

[79] ISO 9241-210, "Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems." 2010.

[80] N. Bevan, "What is the difference between the purpose of usability and user experience evaluation methods," in *Proceedings of the Workshop UXEM*, 2009, vol. 9, pp. 1–4.

[81] N. Bevan, "International standards for usability should be more widely used," *J. Usability Stud.*, vol. 4, no. 3, pp. 106–113, 2009.

[82] H. Miki, "User Experience and Other People: On User Experience Evaluation Framework for Human-Centered Design," in *HCI International 2015 - Posters' Extended Abstracts*, C. Stephanidis, Ed. Springer International Publishing, 2015, pp. 55–59.

[83] ISO/IEC 25022, "Systems and software engineering - Systems and software quality requirements and evaluation (SQuaRE) - Measurement of quality in use." 2016.

[84] C. L. B. Maia and E. S. Furtado, “A Study About Psychophysiological Measures in User Experience Monitoring and Evaluation,” in *Proceedings of the 15th Brazilian Symposium on Human Factors in Computer Systems*, New York, NY, USA, 2016, pp. 7:1–7:9.
[85] ISO/IEC 25000, “Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE.” 2014.
[86] ISO/IEC 25020, “Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Measurement reference model and guide.” 2007.
[87] ISO 9355-2, “Ergonomic requirements for the design of displays and control actuators -- Part 2: Displays.” 1999.
[88] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of psychophysiology*. Cambridge University Press, 2007.
[89] L. E. Nacke, “An introduction to physiological player metrics for evaluating games,” in *Game Analytics*, Springer, 2013, pp. 585–619.
[90] C. L. B. Maia and E. S. Furtado, “A Systematic Review About User Experience Evaluation,” in *Design, User Experience, and Usability: Design Thinking and Methods*, A. Marcus, Ed. Springer International Publishing, 2016, pp. 445–455.
[91] ISO/IEC 25021, “Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality measure elements.” 2012.
[92] R. Bernhaupt, “User Experience Evaluation Methods in the Games Development Life Cycle,” in *Game User Experience Evaluation*, Springer, 2015, pp. 1–8.
[93] C. E. Izard, “Basic emotions, relations among emotions, and emotion-cognition relations.,” 1992.
[94] P. Ekman, “An argument for basic emotions,” *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992.
[95] “CogniFit,” *Cognition and Cognitive Science - The Importance of Cognition*. [Online]. Available: https://www.cognifit.com/cognition. [Accessed: 03-Sep-2018].
[96] “What Is The Definition Of Cognition In Psychology? | Betterhelp.” [Online]. Available: https://www.betterhelp.com/advice/psychologists/what-is-the-definition-of-cognition-in-psychology/. [Accessed: 04-May-2018].
[97] C. Rico-Olarte, D. M. López, and S. Kepplinger, “Towards a Conceptual Framework for the Objective Evaluation of User Experience,” in *International Conference of Design, User Experience, and Usability*, 2018, pp. 546–559.
[98] P. E. Ludwig and S. C. Dulebohn, “Neuroanatomy, Central Nervous System (CNS),” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2018.
[99] K. W. Wills, “Differences Between ‘Physical’ & ‘Physiological,’” *Sciencing*. [Online]. Available: https://sciencing.com/differences-between-physical-physiological-8774303.html. [Accessed: 06-Nov-2017].
[100] D. A. Pizzagalli, “Electroencephalography and high-density electrophysiological source localization,” *Handb. Psychophysiol.*, vol. 3, pp. 56–84, 2007.
[101] L. E. Nacke, “Games User Research and Physiological Game Evaluation,” in *Game User Experience Evaluation*, Springer, Cham, 2015, pp. 63–86.

[102] R. R. Wehbe and L. Nacke, “An Introduction to EEG Analysis Techniques and Brain-Computer Interfaces for Games User Researchers.,” in *DiGRA Conference*, 2013.
[103] M. E. Dawson, A. M. Schell, and D. L. Filion, “The electrodermal system,” *Handb. Psychophysiol.*, vol. 3, pp. 200–223, 2007.
[104] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.
[105] G. G. Berntson, K. S. Quigley, and D. Lozano, “Cardiovascular psychophysiology,” *Handb. Psychophysiol.*, vol. 3, pp. 182–210, 2007.
[106] T. S. Lorig, “The respiratory system,” *Handb. Psychophysiol.*, vol. 3, pp. 243–256, 2007.
[107] M. M. Bradley, M. Codispoti, B. N. Cuthbert, and P. J. Lang, “Emotion and motivation I: defensive and appetitive reactions in picture processing.,” *Emotion*, vol. 1, no. 3, p. 276, 2001.
[108] R. L. Mandryk, K. M. Inkpen, and T. W. Calvert, “Using psychophysiological techniques to measure user experience with entertainment technologies,” *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 141–158, Mar. 2006.
[109] L. G. Tassinary, J. T. Cacioppo, and E. J. Vanman, “The Skeletomotor System: Surface,” *Handb. Psychophysiol.*, p. 267, 2007.
[110] G. Gratton, “Biosignal Processing,” *Handb. Psychophysiol.*, vol. 3, pp. 846–870, 2007.
[111] M. C. Montañés, “Psicología de la emoción: el proceso emocional,” *Univ. Valencia*, 2005.
[112] J. Hernandez, I. Riobo, A. Rozga, G. D. Abowd, and R. W. Picard, “Using electrodermal activity to recognize ease of engagement in children during social interactions,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014, pp. 307–317.
[113] L. Yao *et al.*, “Using Physiological Measures to Evaluate User Experience of Mobile Applications,” in *Engineering Psychology and Cognitive Ergonomics*, 2014, pp. 301–310.
[114] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, “A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments,” *Psychophysiology*, vol. 49, pp. 1017–1034, 2013.
[115] “What should I know to use EDA data in my experiment?,” *Empatica Support*. [Online]. Available: http://support.empatica.com/hc/en-us/articles/203621955-What-should-I-know-to-use-EDA-data-in-my-experiment-. [Accessed: 23-Sep-2018].
[116] M. A. Betancourt, L. S. Dethorne, K. Karahalios, and J. G. Kim, “Skin Conductance As an In Situ Marker for Emotional Arousal in Children with Neurodevelopmental Communication Impairments: Methodological Considerations and Clinical Implications,” *ACM Trans Access Comput*, vol. 9, no. 3, pp. 8:1–8:29, Feb. 2017.
[117] Empatica, “E4 wristband from Empatica - User’s manual.” 2018.
[118] R. Zangróniz, A. Martínez-Rodrigo, J. M. Pastor, M. T. López, and A. Fernández-Caballero, “Electrodermal Activity Sensor for Classification of Calm/Distress Condition,” *Sensors*, vol. 17, no. 10, p. 2324, 2017.

[119] E.-H. Jang, B.-J. Park, S.-H. Kim, Y. Eum, and J.-H. Sohn, “Identification of the optimal emotion recognition algorithm using physiological signals,” in *Engineering and Industries (ICEI), 2011 International Conference on*, 2011, pp. 1–6.

[120] F. Canento, A. Fred, H. Silva, H. Gamboa, and A. Lourenço, “Multimodal biosignal sensor data handling for emotion recognition,” in *Sensors, 2011 IEEE*, 2011, pp. 647–650.

[121] M. Callejas-Cuervo, L. A. Martínez-Tejada, and A. C. Alarcón-Aldana, “Emotion recognition techniques using physiological signals and video games-Systematic review,” *Fac. Ing.*, vol. 26, no. 46, pp. 19–28, 2017.

[122] P. Chapman, J. Clinton, R. Kerber, and T. Khabaza, *CRISP-DM 1.0 Step-by-step data mining guide*. SPSS, 2000.

[123] R. Henriques, A. Paiva, and C. Antunes, “On the Need of New Methods to Mine Electrodermal Activity in Emotion-Centered Studies,” in *Agents and Data Mining Interaction*, 2013, pp. 203–215.

[124] R. Henriques and A. Paiva, “Descriptive Models of Emotion-Learning Useful Abstractions from Physiological Responses during Affective Interactions.,” in *PhyCS*, 2014, pp. 393–400.

[125] S. Taylor, N. Jaques, W. “Vincent” Chen, S. Fedor, A. Sano, and R. W. Picard, “Automatic identification of artifacts in electrodermal activity data,” *2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC*, pp. 1934–1937, 2015.

[126] H. Feng, H. M. Golshan, and M. H. Mahoor, “A Wavelet-based Approach to Emotion Classification using EDA Signals,” *Expert Syst. Appl.*, 2018.

[127] A. Youn, T. Chaspari, R. Gupta, L. I. S. Duker, S. A. Cermak, and S. S. Narayanan, “Capturing the Structure of Electrodermal Activity with Deep Neural Networks.”

[128] B. Kitchenham, S. Linkman, and D. Law, “DESMET: a methodology for evaluating software engineering methods and tools,” *Comput. Control Eng. J.*, vol. 8, no. 3, pp. 120–126, 1997.

[129] N. V. N. Chism, E. Douglas, and W. J. Hilson Jr, “Qualitative research basics: A guide for engineering educators,” *Rigorous Res. Eng. Educ. NSF DUE-0341127*, 2008.

[130] “What is peer review?,” *Elsevier’s Home*. [Online]. Available: https://www.elsevier.com/reviewers/what-is-peer-review. [Accessed: 24-Sep-2018].

[131] C. Godin, F. Prost-Boucle, A. Campagne, S. Charbonnier, S. Bonnet, and A. Vidal, “Selection of the most relevant physiological features for classifying emotion,” *Emotion*, vol. 40, p. 20, 2015.

[132] X. Zhang, Y. Lyu, X. Hu, Z. Hu, Y. Shi, and H. Yin, “Evaluating Photoplethysmogram as a Real-Time Cognitive Load Assessment during Game Playing,” *Int. J. Human–Computer Interact.*, vol. 34, no. 8, pp. 695–706, Aug. 2018.

[133] I. R. Kleckner *et al.*, “Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1460–1467, 2018.

[134] "Medical Definition of MEMORY SPAN." [Online]. Available: https://www.merriam-webster.com/medical/memory+span. [Accessed: 27-Sep-2018].
[135] J. Lopéz, *Dificultades Infantiles en el Aprendizaje*, 1st ed. Grupo Cultural de Colombia LTDA, 2017.
[136] B. Osgood, "The Fourier transform and its applications," *Lect. Notes EE*, vol. 261, p. 20, 2009.
[137] I. Daubechies, *Ten lectures on wavelets*, vol. 61. Siam, 1992.
[138] H. L. Resnikoff and R. O. Wells, "The Mallat Algorithm," in *Wavelet Analysis: The Scalable Structure of Information*, H. L. Resnikoff and R. O. Wells, Eds. New York, NY: Springer New York, 1998, pp. 191–201.
[139] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.
[140] C. Apa, R. Robaina, S. de León, and D. Vallespir, "Conceptos de ingeniería de software empírica," *Rep. Téc. 10-02*, 2010.
[141] R. H. Sampieri, C. F. Collado, and P. B. Lucio, "Metodología de la Investigación," *McGraw-Hill*, 2012.
[142] C. Sammut and G. I. Webb, *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.
[143] J. Brownlee, "Supervised and Unsupervised Machine Learning Algorithms," *Machine Learning Mastery*, 15-Mar-2016. .
[144] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC, 2012.
[145] T. K. Ho, "Random decision forests," in *Document analysis and recognition, 1995., proceedings of the third international conference on*, 1995, vol. 1, pp. 278–282.
[146] J. C. Platt, "12 fast training of support vector machines using sequential minimal optimization," *Adv. Kernel Methods*, pp. 185–208, 1999.
[147] "Parameter estimation using grid search with cross-validation — scikit-learn 0.20.0 documentation." [Online]. Available: http://scikit-learn.org/stable/auto_examples/model_selection/plot_grid_search_digits.html. [Accessed: 28-Sep-2018].
[148] "Multilayer Perceptron," *Skymind*. [Online]. Available: http://skymind.ai/wiki/multilayer-perceptron. [Accessed: 28-Sep-2018].
[149] B. Kamiński, M. Jakubczyk, and P. Szufel, "A framework for sensitivity analysis of decision trees," *Cent. Eur. J. Oper. Res.*, vol. 26, no. 1, pp. 135–159, 2018.
[150] I. Rish and others, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, vol. 3, pp. 41–46.
[151] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992.
[152] D. Corrales, A. Ledezma, and J. Corrales, "From Theory to Practice: A Data Quality Framework for Classification Tasks," *Symmetry*, vol. 10, no. 7, p. 248, 2018.

[153] M. J. Warrens, “New Interpretations of Cohen’s Kappa,” *Journal of Mathematics*, 2014. [Online]. Available: https://www.hindawi.com/journals/jmath/2014/203907/. [Accessed: 28-Sep-2018].

[154] M. L. McHugh, “Interrater reliability: the kappa statistic,” *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, Oct. 2012.

[155] Jojoa, “Definicion de METHONTOLOGY - ¿qué es METHONTOLOGY?” [Online]. Available: https://sites.google.com/site/jojooa/inteligencia-artificial/definicion-de-methontology-que-es-methontology. [Accessed: 14-Nov-2017].

[156] V. Roto and E. Kaasinen, “The second international workshop on mobile internet user experience,” in *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, 2008, pp. 571–573.

[157] Oxford Dictionaries, “Definition of response,” *Oxford Dictionaries | English*. [Online]. Available: https://en.oxforddictionaries.com/definition/response. [Accessed: 06-Nov-2017].

[158] “Los estados internos | PNLnet.” .

[159] “Definición de condición física — Definicion.de,” *Definición.de*. [Online]. Available: https://definicion.de/condicion-fisica/. [Accessed: 03-May-2018].

[160] Y. Fu, H. V. Leong, G. Ngai, and S. Chan, “Non-Intrusive Health-Monitoring Devices,” *Encycl. E-Health Telemed.*, pp. 711–721, 2016.

[161] “What Is ‘Behavior’ Anyway?,” *Psychology Today*. [Online]. Available: http://www.psychologytoday.com/blog/the-prince-evolution/201207/what-is-behavior-anyway. [Accessed: 04-May-2018].

[162] “Definition of LEARNING.” [Online]. Available: https://www.merriam-webster.com/dictionary/learning. [Accessed: 29-Sep-2018].

[163] “Definition of LANGUAGE.” [Online]. Available: https://www.merriam-webster.com/dictionary/language. [Accessed: 29-Sep-2018].

[164] “Reasoning definition and meaning | Collins English Dictionary.” [Online]. Available: https://www.collinsdictionary.com/dictionary/english/reasoning. [Accessed: 29-Sep-2018].

[165] “Definition of DISGUST.” [Online]. Available: https://www.merriam-webster.com/dictionary/disgust. [Accessed: 04-May-2018].

[166] “Definition of PLEASURE.” [Online]. Available: https://www.merriam-webster.com/dictionary/pleasure. [Accessed: 04-May-2018].

Appendix A

Taxonomies from User Experience Concepts

The *Methontology* methodology [155] provides guidelines on how to carry out the development of ontologies, conceptualizing them through a set of intermediate representations, where the representations are concepts, relations, constants, attributes, and rules.

To implement the construction of the conceptual framework, we took into account the first two tasks of the methodology to have a clearer view of the relationships between the concepts studied:

- Task 1: Build the glossary of terms.
  First, we build a glossary of terms that includes all the relevant terms of the domain, their descriptions in natural language, and their acronyms.
- Task 2: Construct concepts taxonomies.
  Once the glossary of terms contains sufficient terms, we construct the taxonomies of concepts that define its hierarchy.

In this annex, we define each of the concepts involved in the UX and the evaluation of the UX from the two perspectives considered: user and system. Then we introduce the taxonomies created.

A.1. Definitions

The conceptual foundation given by the definition of the UX bounded by the ISO serves as a basis to gather the necessary terms, along with the terms from the SQuaRE series of standards having an implication from the UX point of view.

User experience

There are several definitions on what UX is. We stick to the ISO definition, however, here is another one to give clarity: “a term that describes user’s feelings towards a specific product, system, or object during and after interacting with it. Various aspects influence the feelings, such as user’s expectations, the conditions in which the interaction takes place and the system’s ability to serve user’s current needs” [156].

Person/User

A human being interacting with the system, product or object. The experiencing agent/subject of interest [71].

User’s perceptions

Perception is the knowledge about the environment; there are levels of perception: detection, identification, and interpretation. Detection is the perceptual process by which the person becomes aware of the presence of a signal, identification is the perceptual process by which the person distinguished the detected signal from others, and interpretation is the combination of perceptual and cognitive processes, this is where contents and significance of the identified signal are recognized [87].

User’s responses

They are part of a psychophysiological process that occurs in the CNS and constitutes a reply or a reaction due to the application of a stimulus [157]. There are two major types of responses: the physical responses and the psychological responses.

User’s physical responses

An excitation of a nerve impulse caused by a change or event. A physical reaction (body response) to a specific stimulus or situation [157].

User's psychological responses

Psychological responses are the phenomena related to the cognition, emotion, experience and behavior of organisms due to stimulus in the physical and social environment [88].

User's internal state

The overall subjective experience of a person at a particular time. This experience arises as a result of the person's interpretation of a situation, whether it be an internal experience or an external event [158].

User's physical state

State of the body nature (physical condition) in which someone is or the external constitution of a person [159].

Psychophysiology

"The scientific study of social, psychological, and behavioral phenomena as related to and revealed through physiological principles and events in functional organisms" [88, p. 4].

Physiological signals

To understand what the physiological responses are, we need to check on the big picture: what is called physiological refers to the "functions" in the body [99]. The physiological signals are readings or measurements produced by the physiological process of human beings [160].

- Electroencephalography (EEG)
  Measure of brainwave activity of the human body [89, p. 595].
- Electrodermal activity (EDA)
  It measures changes in the passive electrical conductivity of the skin relating to increases or decreases in sweat gland activity [89, p. 601].
- Cardiovascular physiology
  There are many cardiovascular measures: the most common signals are electrocardiography (ECG), heart rate (HR), interbeat interval (IBI), heart rate variability (HRV), blood volume pulse (BVP), and blood pressure (BP) [89, p. 603].

- Gastrointestinal response
  Electrogastrography measures the gastrointestinal system and refers to the recording of electrogastrograms (EGGs). These EGGs reflect the gastric myoelectrical activity recorded from the abdominal surface with electrodes placed on the skin [88, p. 211].
- Sexual response
  Gender determines the type of measurement of the genital response to know sexual arousal. For men, it is common to measure the volume, circumference, and stiffness of the penis. In contrast, in women, oxygenation and thermal elimination, temperature, and the amplitude of the vaginal and labial pulse are measured [88, p. 254].
- Respiratory system
  Respiration can be measured as the rate or volume at which an individual exchanges air in their lungs [76, p. 218].
- Electromyography (EMG)
  Electromyography is the measure of muscle activity through needles or surface electrodes [76, p. 219].

Psychological events
Same as *user's psychological responses*.

Behavior
Observable activity of an organism; anything an organism does that involves action and/or response to stimulation [161].

Cognition
The ability to process information through perception, knowledge acquired through experience, and the subjective characteristics that allow the human being to integrate all of this information to evaluate and interpret the world [95].

- Learning
  "Knowledge or skill acquired by instruction or study, or modification of a behavioral tendency by experience" [162]
- Attention
  Neurological disposition of the brain for the reception of stimuli [135].

- Memory
  The process of replicating or remembering what has been learned and retained [134].
- Language
  Mean of communicating ideas or feelings by using conventionalized signs, sounds, gestures, or marks with understood meanings [163].
- Reasoning
  The process to reach a conclusion after considering all the facts [164].

Emotions

They are a multidimensional experience with at least three response systems: cognitive/subjective, behavioral/expressive, and physiological/adaptive. Any psychological event involves an emotional experience of greater or lesser intensity and quality. The emotional reaction is something omnipresent to every psychological event [111]. There are what some authors call basic emotions [93], [94], such as:

- Happiness
  Pleasurable and desirable state, sense of well-being.
- Anger
  Sensation of energy and impulsiveness, need to act intensively and immediately (physically or verbally) to actively solve the problematic situation.
- Fear
  Concern, fear for one's safety or health; feeling of loss of control.
- Sadness
  Discouragement, melancholy, discouragement, loss of energy.
- Surprise
  Transitional state. Sensation of uncertainty for what is going to happen [111].
- Disgust
  Marked aversion aroused by something highly distasteful [165].
- Pleasure
  A source of delight or joy; a state of gratification [166].
- Contempt
  Need to avoid or move away from the stimulus.

Interactive system

"Combination of interacting elements organized to achieve one or more stated purposes. A system may be considered as a product or as the services it provides" [85].

System quality

"Capability of system product to satisfy stated and implied needs when used under specified conditions".

Process quality

"Ability of a process to satisfy stated and implied stakeholder needs when used in a specified context".

External quality

"Degree to which a system enables its behavior to satisfy stated and implied needs for the system including the software to be used under specified conditions".

Internal quality

"Degree to which a set of static attributes (software architecture, structure and its components) of a system satisfies stated and implied needs for the software to be used under specified conditions".

Quality in-use

"Degree to which a system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, freedom from risk and satisfaction in specific contexts of use". [85]

Effectiveness

"Accuracy and completeness with which users achieve specified goals".

Efficiency

"Resources expended in relation to the accuracy and completeness with which users achieve goals".

Satisfaction

"Degree to which user needs are satisfied when a product or system is used in a specified context of use".

Freedom from risk

"Degree to which the quality of a product or system mitigates or avoids potential risks to economic status, human life, health, or the environment".

Context coverage

"Degree to which a product or system can be used with effectiveness, efficiency, satisfaction, and freedom from risk in both specified contexts of use and in contexts beyond those initially explicitly identified".

Usefulness

"Degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use".

Trust

"Degree to which a user or other stakeholder has confidence that a product or system will behave as intended".

Pleasure

"Degree to which user needs for pleasure are satisfied".

Comfort

"Degree to which user needs for physical comfort are satisfied". [83]

A.2. Taxonomies

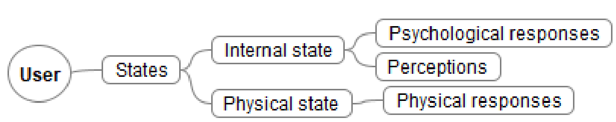

Figure 25. User's taxonomy

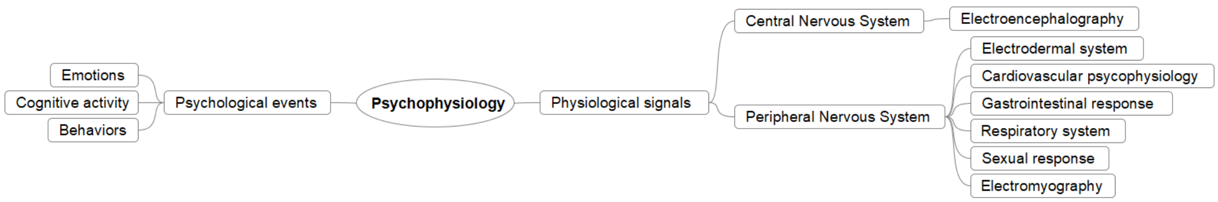

Figure 26. Psychophysiology's taxonomy

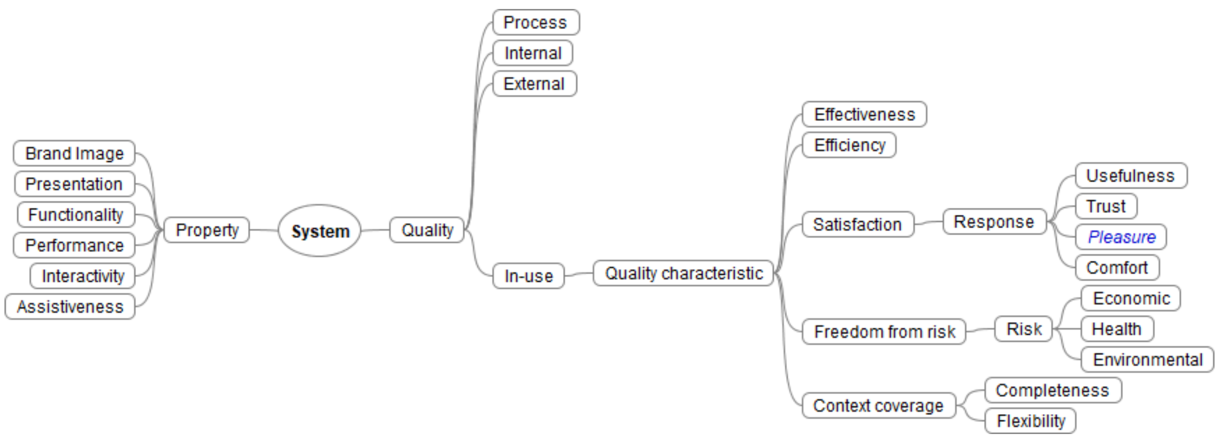

Figure 27. System's taxonomy

Appendix B

Informed Consent

Proyecto:
HapHop Fisio II - Un Juego Serio Interactivo basado en Movimiento (Exergame) para el Soporte a Terapias Cognitivas en Niños.

Estimados Padres y Madres de Familia:

Fisiocenter S.A.S y la Universidad del Cauca apoyan la práctica de obtener el consentimiento informado de, y proteger a, los sujetos humanos que participen en investigaciones. La siguiente información tiene por objeto ayudarle a decidir si permitirá que su hijo(a) participe en el presente estudio. Su(s) hijo(s) está(n) cordialmente invitados a formar parte de este proyecto de investigación haciendo uso del juego serio e interactivo HapHop Fisio II.

El objetivo es desarrollar la segunda etapa del juego serio interactivo HapHop Fisio, el cual está basado en movimiento (ExerGame) para el soporte a las terapias cognitivas en niños, con el fin de potencializar su aprendizaje. El proyecto inicia haciendo un análisis de las necesidades de los niños y los profesionales en terapias cognitivas para definir las adaptaciones necesarias para la nueva etapa del juego, posteriormente viene la etapa de construcción del prototipo y la evaluación de su funcionalidad e interacción.

Nos gustaría que su hijo(a) participara, pero tal participación es estrictamente voluntaria. Le aseguramos que el nombre de su hijo(a) no se mencionara en los hallazgos de la investigación. La información solo se identificará mediante un código numérico.

Usted no recibirá incentivo de índole económico (no se le dará dinero) por participar en el estudio, pero los resultados del desempeño de su hijo durante la actividad le serán compartidos si son de su interés.

Si le gustaría tener información adicional acerca del estudio antes o después que finalice, puede ponerse en contacto con cualquiera de nosotros por teléfono o correo. Agradecemos su atención y apreciamos su interés y cooperación.

La información sobre su hijo(a) será vista únicamente por nuestro grupo de investigación. La información (fotos, videos y cuestionarios) será guardada en cajón bajo llave en la oficina del profesor Diego Mauricio López Gutiérrez en la facultad de Ingeniería Electrónica de la Universidad del Cauca. Nadie fuera de nuestro equipo de investigación vera la información sobre su hijo(a).

Solo el equipo de investigación sabrá cuál es el código único de su hijo(a). De este modo, podremos comunicarle si encontramos algo que requiera su atención con respecto al proceso de aprendizaje que adelanta el (la) niño(a). Cuando expliquemos nuestra investigación a otras personas, no usaremos su nombre o nada que permita que otra gente sepa quién es usted o su hijo(a). La información (fotos, videos y cuestionarios) será guardada cumpliendo los criterios de confidencialidad y respeto. Cabe resaltar que todos los datos personales utilizados en este proyecto solo se manejaran en el mismo y no serán utilizados en otras investigaciones.

Se construirá una base de datos, la cual estará en un computador del grupo de investigación, en custodia del investigador principal, Diego Mauricio López Gutiérrez y a esta información solo tendrán acceso los aliados e investigadores del proyecto.

Los resultados serán divulgados a través de medios de comunicación y eventos académicos, sin embargo, no se mencionaran los nombres de los niños y niñas, de modo que otras personas puedan aprender de nuestra investigación, sin conocer su identidad. La información obtenida será de utilidad para la decisión acerca de la manera de mejorar la salud de niños con discapacidad cognitiva.

Usted tendrá derecho a retirar a su hijo(a) cuando lo desee, su participación es voluntaria, y en caso de retirarse los datos de su hijo(a) serán eliminados. Esta investigación contiene los elementos éticos que la ley y la doctrina exigen (código Helsinki – código Nuremberg según resolución 8430 de 1993) que rigen la ética en la investigación científica en Colombia. Se garantiza total confidencialidad con los datos recolectados.

Si tiene preguntas, puede contactar al docente Diego Mauricio López Gutiérrez, en la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca. Teléfono 8209800 ext. 2175 o al correo dmlopez@unicauca.edu.co.

Esta propuesta ha sido evaluada y aprobada por el comité de ética de la Vicerrectoría de Investigaciones de la Universidad del Cauca, Popayán.

Cláusulas estándar:

- Mi decisión de participar en el estudio es completamente voluntaria.
- Los procedimientos principales, han sido expuestos en un lenguaje comprensible, logrando total entendimiento del mismo.
- La información obtenida en este estudio, será mantenida con estricta confidencialidad por el grupo de investigadores, y solo para el propósito del protocolo descrito en este documento.
- Me han explicado los beneficios de este estudio.
- Me han explicado que me podré retirar en cualquier momento del estudio, sin que ello acarree perjuicio en este estudio.
- Los resultados de este estudio pueden ser divulgados en eventos nacionales y/o internacionales o ser publicados en revistas científicas sin identificar el nombre de mi(s) hijo(a)(s). Además tendré derecho a conocer los resultados. Si tengo una pregunta durante o después del procedimiento puedo contactar a los investigadores.

Con relación al proceso de rehabilitación. Acepto que se realicé a mi hijo los siguientes procedimientos:

- Valoración el estado funcional del paciente, descripción de deficiencias estructurales y funcionales, discapacidades encontradas y riesgos de deficiencias secundarias.
- Evaluación y diagnóstico de aprendizaje, para confirmar las posibilidades terapéuticas, determinar el potencial de recuperación y proponer áreas de intervención del tratamiento.
- Presentación de estímulos visuales, auditivos y propioceptivos que estimules canales de recepción del aprendizaje.
- Ejecución de ejercicios para mejorar dispositivos básicos del aprendizaje y habilidades específicas de lectura y escritura.

Informo que entiendo que según el diagnóstico durante la terapia puede haber algunos riesgos, tales como:

Alteración en el ritmo respiratorio y la frecuencia cardiaca

- Dado que el paciente pueda presentar alteraciones del equilibrio y/o conducta, es posible que durante sus desplazamientos puedan ocurrir caídas o algún tipo de accidente como consecuencia de su condición.
- Fatiga muscular, enrojecimiento de la piel, malestar general.
- Afirmo que me han permitido aclarar dudas y me han dado la información necesaria para el tratamiento a recibir.
- Afirmo que me han informado que el potencial rehabilitatorio y la misma evolución del paciente tienen un límite, por lo cual la funcionalidad puede ser menor al esperado para su edad cronológica.
- Con mi firma dejo constancia de haber sido informado el programa de atención y doy consentimiento para que se realicen los procedimientos respectivos, aceptando los riesgos y complicaciones inherentes a él.
- Finalmente se me ha informado de mi derecho a rechazar del tratamiento o revocar este consentimiento y de las consecuencias de esta decisión.

Certificado de consentimiento

Nombre del participante: ____

Nombre legible del padre, madre o tutor: _________________________________

Firma del padre, madre o tutor: __________________________ Fecha: ________

He leído exactamente o he sido testigo de la lectura correcta del consentimiento al participante potencial, y éste ha tenido la posibilidad de hacer preguntas. Confirmo que el participante ha dado consentimiento libremente.

Nombre legible del investigador: Diego Mauricio López Gutiérrez.

Firma del Investigador: ____

Se da una copia de éste consentimiento informado al participante